

Where next on anti-proliferation?

The successful conclusion of last week's review conference of the NPT, and the indefinite prolongation of the treaty, should not blind us to the need for action before complacency overtakes us.

If China had not carried out an underground nuclear explosion at the beginning of this week, it would have been entirely reasonable to be throwing hats in the air to celebrate last week's successful conclusion of the review conference of the Non-Proliferation Treaty (NPT). Against expectation, the conference agreed to an indefinite extension of the treaty. There will still be review conferences every five years, but none of those will be an occasion when the treaty will collapse automatically unless there is an affirmative vote in favour of continuation. This is good news in the sense that we shall all now be able to sleep more easily at night. But we probably shall. China's explosion should remind us how unwise that would be.

Last week's meeting in no sense implies that non-proliferation has now been signed and sealed. Recent experience, with North Korea and Iraq, for example, is a vivid reminder that governments' nuclear inclinations are not easily predicted and can seem to pop out of nowhere. Those black sheep will not be the last in the rest of time. Moreover, there remain the threats to peace of mind, and possibly to peace, presented by Israel, India and Pakistan, all of which have deliberately advertised nuclear inclinations and which may also be nuclear powers on some undetermined scale. Those problems will not melt away of their own accord.

Where to start? The answer is with China, which has let it be known during the past month's negotiations in New York that it would not rock the boat on proliferation if three conditions were satisfied: a comprehensive test-ban comprehensively adhered to, a declaration by nuclear powers that they would not attack non-nuclear powers with nuclear weapons and a declaration by nuclear powers that they would never be the first to use nuclear weapons in a conflict. The first condition makes sense because it is believable. The other two do not, because they are predicated on trust between potential adversaries.

Luckily, there are constructive steps that can be taken. Signing and then ratifying a comprehensive test-ban treaty within a year or so should be a first objective. Persuading China that there are other and more reliable ways of preventing nuclear attacks than public declarations that there will be none should be the second. A verifiable cut-off of the production of missile material for military purposes would powerfully inhibit the stockpiling of nuclear weapons and, ultimately, would restrain the capacity of existing nuclear powers to threaten smaller fry with destruction. For, in a more open international community, public knowledge of

who has what will be an effective modulator of what now seems increasingly old-fashioned power politics.

It will be a great misfortune if, otherwise, the years ahead are destined to be like the past twenty-five. Inspectors repeatedly trudging off to look at strange reactors, and occasionally suspecting that something is amiss, will not put the fear of proliferation to rest. Nor will declarations of no first-use. The best hope is that the nuclear powers will wake up to the truth that greater restraint on their part, agreed upon collectively, is the only way of damping down the ambitions of now non-nuclear states. Last week's successful ending should not send us all to sleep. □

AIDS pathology unknown

HIV infection provokes hyperactivity of the immune system, but the causes of that are far from understood.

THE clutch of contributions to Scientific Correspondence (page 193) this week deserves a reading, both for its inherent interest and for what it says about the present state of AIDS research. It will be recalled this journal published in January an account of research that showed that the infection of a person by the virus HIV ordinarily evokes not the previously suspected quiescence of the immune system, but a rapid turnover both of the vulnerable lymphocytes and of the virus itself. The then-general opinion that the first reaction of the human body to infection by HIV is a kind of indifference was dramatically and directly challenged. Nothing that has since come to light denies the challenge. But it has also become plain that too little is yet known of the dynamics of the immune system. That is a gap to fill.

The second arresting feature of this correspondence is the letter from Dr Peter Duesberg and his colleague, Dr Harvey Bialy, which has been published without change. Sadly, there seems no way in which the authors concerned can be persuaded that "free and fair scientific debate" is ordinarily understood to mean a progressive process, one in which each of two sides learns from what the other says. A restatement of earlier and well-known positions is not that at all. On this occasion, Duesberg and Bialy's citation of Loveday in their cause is especially inappropriate, given Loveday's name among the authors of a letter supporting Wei *et al.* and Ho *et al.* But no further solicitation of Duesberg's opinion is called for. □



NIH loses long-standing protection from impact of heavy budget cuts

Washington. The Senate budget committee has agreed a budget resolution which implies cuts of at least \$1.1 billion next year from the \$11.3 billion budget of the National Institutes of Health (NIH). The move has stunned lobbyists who had thought that biomedical research was relatively safe from the budget cutting fever sweeping through Washington.

As the Senate budget committee, chaired by Peter Domenici (Republican, New Mexico), released a proposal for cuts of \$3 billion from discretionary health programmes — of which NIH is the largest component — NIH officials estimated that it would require a cut of \$1.8 billion (16 per cent) in their budget.

But committee staff said that the budget assumed a reduction of only 10 per cent in NIH funding from this year's level for the 1996 financial year, which begins 1 October.

The proposal sent shock waves through the biomedical research community. It had hoped to avoid the fate of scientists who depend on agencies such as the Department of Energy and the National Aeronautics and Space Administration (NASA) which are regarded as being more politically vulnerable than NIH.

The proposed cuts would be "a major catastrophe for basic research in this country," says George Mandel, chair of the National Caucus of Basic Biomedical Science Chairs, and head of the pharmacology department at George Washington University, Washington DC.

Before it is passed, the Senate budget resolution must be approved by the Senate and reconciled with the House version, crafted by budget committee chairman John Kasich (Republican, Ohio).

The House resolution proposed a \$500

million (5 per cent) cut for NIH, and level funding for five years after that, although officials say that other pending House provisions add up to a total cut of 9 per cent in NIH funding next year. Neither proposal allows for inflation, which could consume a further 4 per cent from NIH's budget.

In September, appropriations committees on both sides will allocate budget funds on the basis of the limits set by the agreed budget resolution. President Clinton is due to sign — or veto — their budget by October 1.

As far as other agencies are concerned, the Republican resolutions include a proposal from the House side to cut NASA's budget from \$14.5 billion this year to \$13.7 billion next year, and to \$11.7 billion by 2000. Over five years, \$2.7 billion would come from the space agency's Mission to Planet Earth programme and another \$1.5 billion from privatizing the space shuttle.

This proposal would maintain space station funding at \$2.1 billion a year. But George Brown (Democrat, California), the senior Democrat on the House science committee, says that this would lead to unsustainable pressures on other parts of NASA.

"The Republican budget calls into question the possibility of achieving any reasonable balance among NASA programmes," says Brown. After the budget was announced, conflict between Brown and Bob Walker (Republican, Pennsylvania), chair of the science committee, over this issue led to the abandonment of their bi-partisan approach to authorization legislation for NASA.

The House resolution also proposes abolishing the Department of Energy, without saying what should happen to its constituent parts. Walker, as vice-chair of the budget committee, says he tried and failed to incorporate his plan for a new Department of Science in the resolution.

The House bill includes steep reductions in research and development on energy supply, from \$3.3 billion this year to \$2.6 billion next. But within that, it offers protection both for basic science and for the administration's Scientific Facilities Initiative.

The National Science Foundation (NSF) fares better than other science agencies under both Republican plans. The Senate proposes a cut of \$100 million from the

agency's \$3 billion budget this year, but the House allows 3 per cent annual growth for NSF research programmes, except for the social sciences, a move which could lead to the complete elimination of the NSF's \$120 million social sciences programme.

Republicans in both House and Senate propose ending the Advanced Technology Programme (ATP), a scheme for supporting industrial technology which has had the misfortune to be seen as a symbol of President Clinton's science and technology policy.

The House resolution cuts funding for the United States Geological Survey (USGS) by one-fifth, and abolishes the National Biological Service (NBS). Both House and Senate would abolish the Office of Technology Assessment, which provides Congress with impartial, if long-winded, advice on science and technology issues.

Walker, whose science committee has jurisdiction over most civil science programmes except NIH, says that the budget will maintain level funding for basic research, at the expense of applied research and development. "We went after corporate welfare," he says.

Justifying the cuts on NASA's Mission to Planet Earth programme, he says that it "was structured to take care of political interests" in its orientation toward global warming. Although the programme has a scientific mission, the money used to build its satellites and data gathering systems is categorized as 'development' rather than 'research'.

But Brown scorns Walker's assertion that the budget proposal protect basic science, saying that it "threatens to destroy the investment portfolio of this nation". His staff estimate that, if adopted, it will reduce science committee programmes by 35 per cent by the year 2000.

As for NIH, the bulk of any cuts will come from the roughly \$9 billion of its \$11.3 billion budget allocated to medical schools and other outside research centres. Sam Silverstein, president of the Federation of American Societies for Experimental Biology, and head of physiology at Columbia University, New York, brands the Senate proposal as "a disaster, which will lead to the destruction of the system" for funding biomedical research in the United States.

Silverstein and Mandel both suggest that the cut could virtually eliminate new NIH grants next year; Silverstein says it could mean a 1 per cent success rate for them. Alternatively, both say, the NIH might have to cut off funding for research projects already underway, in order to free up money for new work.

Colin Macilwain



Domenici: seeks \$3bn off health spending.

US campaign to target patenting on life

Washington. A broad coalition of religious leaders from more than 80 faiths and denominations is due to launch a campaign in Washington today (18 May) against the patenting of human genes, cells and organs, as well as genetically engineered animals.

The campaign has been organized by the Foundation on Economic Trends together with the General Board of Church and Society of the United Methodist Church (see *Nature* 374, 103; 1995). It is intended to raise "critical theological concerns over the patenting of life" in churches, synagogues, mosques and temples across the United States. □

Japan is first non-CERN contributor to LHC

Tokyo. Partly as a result of the Kobe earthquake, Japan's Ministry of Education, Science and Culture (Monbusho) has managed to surprise the world of high-energy physics with a quick decision to contribute ¥5 billion (US\$60 million) towards the construction of Europe's Large Hadron Collider (LHC). Japan is the first foreign country not belonging to the European Laboratory for Particle Physics (CERN) to back the LHC.

The contribution in cash will be drawn from a supplementary budget of ¥2,726 billion drawn up by the government in response to the Kobe disaster. The budget, which was approved by the cabinet last week and is expected to receive quick approval by Japan's parliament, the Diet, is primarily intended to finance reconstruction of the Kobe region and disaster prevention activities. But it also contains ¥320 billion for information networks and science — including the budget for the LHC.

Christopher Llewellyn Smith, director general of CERN, said on Monday that he is "absolutely delighted" by the Japanese decision. He said he had been confident of a Japanese contribution after visiting Tokyo in March (see *Nature* 374, 103; 1995), but has been pleasantly surprised with both the speed at which the money has been found and its amount.

According to Llewellyn Smith, Japan's contribution is not itself sufficient to bring forward LHC's construction schedule. But he says it is a "great step forward".

The LHC is scheduled to reach its maxi-

mum energy of 14 TeV by 2008 in two phases of construction. But if CERN can raise about 600 million Swiss francs in non-member contributions — about ten times the promised contribution from Japan — construction could be completed in one phase by 2004.

Japan is prepared to consider further contributions "in the light of the development of cooperation between CERN and Monbusho, the progress of the LHC project, and contributions from other non-member states", says a spokesman for the ministry. Kaoru Yosano, the minister of education, says that Japan took the lead in contributing in the hope that it "will accelerate international cooperation in the project".

But the United States has given no commitment so far that it will follow in Japan's footsteps. US high-energy physicists remain keen to support construction of the LHC following the demise of the Superconducting Super Collider, and the US Department of Energy (DoE) has expressed interest in supporting the LHC. But the department faces an uphill battle with Congress over its budget, and a US decision to enter formal negotiations on the LHC is not

expected to be made until later in the year.

Shuji Orito, head of Tokyo University's International Center for Elementary Particle Physics, which participates in experiments on CERN's electron-positron collider LEP, describes the Japanese decision as a "great initiative" by Monbusho.

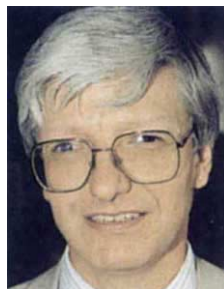
Hirotaka Sugawara, director general of the National Laboratory of High Energy Physics (KEK) in Tsukuba science city, is slightly more cautious, pointing out that there are "many problems that need to be solved". In particular, he says that Japan is hoping for observer status both on CERN's council and on other committees making decisions concerning the LHC.

A decision on this has not yet been made by CERN's council. But Llewellyn Smith is confident that a request for observer status would be looked on "very favourably", and is keen that Japan should consider some sort of new "associate status" to give it a "real voice" in CERN decisions.

KEK is developing some of the dipole magnets for the LHC. Following the funding decision, Sugawara says he expects his laboratory to work also on the quadrupole magnets for fine control of the proton beams.

But he wants European scientists to go to KEK and develop the magnets jointly. Japan, he says, "cannot be held responsible" for operation of the magnets at CERN. He is also keen that CERN should hire some Japanese scientists; until now, CERN has drawn its employees only from member states.

David Swinbanks



Llewellyn Smith: 'absolutely delighted'.

Agency halts HIV survey in dispute over newborn data

Washington. In a move that has surprised some of their closest advisers, the US Centers for Disease Control and Prevention (CDC) in Atlanta have suspended a \$10.5-million surveillance study designed to track HIV infection in women, to avoid being forced by Congress to collect full details of HIV exposure in newborn children.

At a congressional hearing in Washington last week, Helene Gayle, acting director of CDC's national centre for prevention services, announced that Philip Lee, the assistant secretary for health, was asking all states to suspend such testing of newborns.

The hearing was held to discuss a bill introduced by Gary Ackerman (Democrat, New York) requiring states to disclose the results of HIV tests to the legal guardians of babies tested this way. CDC's move has pulled the rug out from under the proposed legislation. But it has also prompted Ackerman to consider introducing a new bill requiring HIV testing for all US newborns.

CDC began its survey of HIV in child-bearing women in 1988 as part of its efforts to track HIV infection in women by testing

the blood of their newborns. With CDC's support, 45 states have been routinely testing samples of blood from newborns for HIV-1 antibodies, removing personal identifying information but retaining limited demographic data. The survey had no termination date, although its funding would have had to be renewed at the end of this year.

Ackerman introduced his legislation to lift the restrictions on the data being reported to CDC, and to require that the legal guardians of all babies testing positively for HIV be notified and counselled. His goal is to enable prospective adoptive parents access to information on the HIV status of babies, and allow parents with infected babies to begin immediate treatment.

But although only 15 to 30 per cent of all babies born to infected mothers are themselves infected, virtually all test positive at birth, because of the presence of maternal antibodies to HIV in the bloodstream.

Suspending the study was "a bold move by CDC — they were boxed into a corner by the Ackerman bill," says Laura Morrison of the Treatment Action Group (TAG). "It was

a compassionate decision in the interest of protecting the privacy of women being screened. 'Unblinding' the study would in effect make testing mandatory."

During the congressional hearing, Gayle pointed out that a recent study by the National Institutes of Health showed that maternal transmission of HIV can be drastically reduced through the administration of AZT during pregnancy.

Gayle said that CDC would like to see more infected women identified, counselled and treated before or during pregnancy, in order to reduce the risk of transmission to unborn infants. Newborn screening, says CDC, is too little too late to prevent HIV transmission to children.

New CDC guidelines will be published next month calling for greater emphasis on prevention by encouraging counselling and voluntary screening of all pregnant women for HIV infection. A panel of experts is being set up to examine ways of promoting the new guidelines for prevention and treatment, and will evaluate the role of the HIV survey in that effort.

Nancy Touchette

Germany seeks to re-balance funding for research centres

Munich. Tensions over the reunification of Germany have fanned disagreements between the country's 16 *Länder* about how much each should pay towards research institutes in their region and what proportion of the costs of each institute should be shared jointly.

The balance of expenditure has been distorted by the effects of reunification, and the situation is being reassessed in the light of the fact that a temporary programme supporting the integration of east German research into the national research effort expires next year. The issue is important because a host *Land* that pays a significant proportion of an institute's budget can exert a corresponding influence over its research programmes.

There are three types of non-university basic research establishments in Germany, each funded in different proportions, federally and locally. The *Bund-Land Kommission* for Education Planning and Research Funding (BLK), the body that brokers political and financial agreements between federal and *Länder* governments, has been asked by the prime ministers of the *Länder* to propose new formulae for dividing funding responsibility between the two sides.

The budgets of the 16 large national research centres are divided 90:10 between federal government and the host *Land*. This formula is not under scrutiny, as it benefits each *Länder* to the same extent, and there is therefore no political reason why the centres should not be relatively evenly distributed.

The funding of the 68 Max Planck institutes is more problematic. Each institute receives half its money from the federal government, and half from the *Länder*. The latter share is further subdivided, with the host *Land* paying a quarter and the rest coming from a central fund, controlled by the Max Planck Society (MPG), to which all *Länder* contribute.

This formula enables the MPG to decide where its institutes should be placed, and what type of research each should do, on scientific grounds alone. But the result is that *Länder* with relatively few Max Planck institutes find themselves providing substantial support for institutes that have been established in other *Länder*.

Rheinland-Pfalz, for example, which has only two institutes, has taken the lead in arguing that the host *Land* should pay much more, perhaps even all of the *Länder* share. In contrast, Bavaria, with several institutes, strongly supports the status quo.

Even more problematic are the other publicly-funded basic research institutes, known collectively as Blue List as they were

originally listed on blue paper. These also receive half their funding from the federal government, but have no general formula for distributing the *Länder* share.

In principle, the host *Land* pays the full 50 per cent. In most cases in western Germany, part of this is reimbursed from a communal pool intended to help even out local expenditure on research between *Länder*. But so far the many Blue List institutes in the new *Länder* — created primarily as a way of continuing to support good science in the Academy of Sciences in the former East Germany — have not benefited from the same reimbursement mechanism.

Yet because there are so many of them — after reunification the number of Blue List institutes jumped from 48 to 82 — most of the old *Länder* want to make sure that the new eastern *Länder* do not receive such reimbursement in future either.

Because of this sudden growth, a new umbrella group for all Blue List institutes, whose total budget is now similar to that of the MPG, was created in 1992. The so-called AG-BL would like to be given similar control over the finances of its institutes as the MPG, but at present this remains a distant dream.

The new *Länder* themselves are keenly aware of the disadvantage of being excluded from the reimbursement mechanism of the west, particularly as their research is heavily locked into the Blue List system. (Only three relatively small national research centres were established in the east, while the MPG prefers not to take over existing research teams, and is only slowly founding institutes.)

But only three of the smallest western *Länder* — Berlin, Bremen and Schleswig-Holstein — feel that the heavy costs of Blue List institutes in the new *Länder* should even be shared by the reimbursement method. Richer and more powerful old *Länder*, such as Bavaria, Baden-Württemberg and Nordrhein-Westfalen, oppose the idea, as they would have to bear most of the subsidy.

Michael Lankeit, a spokesman for the AG-BL, says the issue of who should pay what must be resolved soon to enable the organization to start taking over new institutes, and perhaps closing others. (Germany's science council, the *Wissenschaftsrat*, last year recommended a further eight institutes be added to the Blue List, but so far no action has been taken.)

But he is pessimistic about an agreement being reached by September, as requested by the *Länder* prime ministers. "The sides are too far apart," he says. **Alison Abbott**

US agencies bury hatchet over aid to Russian science

Washington. After nearly three years of delay, and heated wrangling between federal agencies, the United States last week agreed to provide \$5 million to help set up a fund for collaborative research between US scientists and researchers in the former Soviet Union (FSU).

The grant was announced by President Bill Clinton during a Moscow summit meeting with the Russian President, Boris Yeltsin. Money from a defense conversion fund taken out of the budget of the Department of Defense will be matched by the philanthropist George Soros.

The two sums will provide a total of \$10 million to create a Civilian Research and Development Foundation. The Russian government has promised another \$10 million to the fund, and other former Soviet states are expected to provide additional resources.

Congress authorized the creation of the foundation in 1992, as part of its effort to redirect military funding to civilian projects, at the insistence of George Brown (Democrat, California), the former chairman of the House of Representatives Science committee.

But the Defense Department stubbornly refused to release the funds to get it off the ground. Its lawyers argued that they were not legally entitled to release any money unless it was matched by another federal agency (see *Nature* 372, 717;1994). Soros, equally stubborn, had initially vowed not to give any more money to support Russian science unless it was matched by the US government, but his new move is said to have broken the log-jam by giving the money to a National Science Foundation (NSF) trust fund, allowing NSF to become the matching agency.

Now that both sides have relented, the NSF will establish the foundation as a private, nonprofit organization with an independent board of directors. Tom Owens of the NSF's international programmes division says he expects the board to be in place by the end of next month. The first task of the board will be to work with Russian and other FSU science agencies on deciding which areas of research should be funded.

Owens says that "the overwhelming proportion" of funds will be spent in the former Soviet Union. Although each project will have US collaborators, these would generally receive only travel costs and other "marginal" expenses. The foundation's funds will be split evenly between support for basic science projects and for technology projects with potential commercial applications. Tony Reichhardt

Academy under fire on 'wetlands' definition

Washington. In a cautionary twist to the debate about putting federal regulations on a more 'scientific' footing, a US National Academy of Sciences study on wetlands has come under fire from congressional sponsors of a bill whose scientific foundation the study calls into question.

Bud Shuster (Republican, Pennsylvania) and other authors of the Clean Water bill, which the House of Representatives was expected to pass this week, have accused the academy of "playing politics" by releasing its wetlands report a day before debate on the bill.

The academy report does not mention specific legislation. But it does call for a much broader definition of wetlands than is included in the current version of the bill.

At stake is the fate of the nation's swamps, bogs and other wetlands, which are protected by the federal government. Seventy-five per cent of wetlands in the lower 48 states are privately owned, and Shuster and his allies want to narrow the definition of a wetland, removing federal protection from an estimated 50 per cent of existing wetlands so they can be developed for other purposes.

The legal definition of a wetland has been controversial for nearly a decade. Federal agencies responsible for delineating wetlands still follow guidelines prepared in 1987 by the Army Corps of Engineers. Two attempts to rewrite the delineation manual, in 1989 and again in 1991, failed because of the political heat they caused.

The Shuster bill borrows language from the 1991 draft, which environmentalists criticized for excluding too many wetlands. One of its criteria for defining a wetland is that water should be present at the surface for 21 consecutive days during the growing season. Three factors would also each have to be present: wetlands hydrology, characteristic soils and characteristic vegetation.

But the academy study says that the 21-day threshold is not a useful criterion, and would exclude many bona fide wetlands. It also says that in many cases soil and vegetation are enough to identify a wetland without resorting to hydrology.

Even more controversially, the report clearly states that regions with permafrost in Alaska and agricultural wetlands — both of which are political hot potatoes — are in fact wetlands. Furthermore, even though it finds that the government's manual on delineating wetland needs to be improved, it contends that current federal wetlands policies are generally based on sound science.

That conclusion has upset property-rights advocates in Congress, who base much of their argument on claims that the government has over-regulated wetlands. During the House debate, Shuster and other spon-

sors of the bill went on the attack.

"The American people should know that that study was funded by the [Environmental Protection Agency (EPA)] bureaucrats downtown," said Shuster. (Congress ordered the wetlands study in 1992 and authorized EPA to pay for it, as is common practice for federal agencies that require the academy's outside advice.) "So, sadly, the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Wet, wet, wet: the scientific basis of the characterization of wetlands has been questioned by a bill in Congress.

National Academy of Sciences has been politicized for this debate."

An editorial in the conservative *Washington Times* was even more blunt. "A swamp is a swamp," said the newspaper. "The real ones are in no danger from the Shuster bill, no matter what political scientists may say." That prompted some moderate Republicans, including Sherwood Boehlert of New York, along with George Brown of California, ranking Democrat on the House Science Committee, to draft a letter to the newspaper defending the academy's integrity.

William Lewis of the University of Colorado, who chaired the wetlands study, says that those who are trying to read significance into the timing of its release "are mistaken". He points out that the report took more than two years to compile, and that it was the academy's internal review process

— not the legislative calendar — that determined its release date.

But politics did play a role. Boehlert, Gilchrest and other opponents of the Shuster bill had attempted to get House Republican leaders to delay debate on the bill until the academy study was released.

When they were refused, and it appeared the report would not come out in time, they urged the academy to speed up its procedures, and the release was moved forward from 18 May to 9 May. But, laments a Boehlert staff member who would have liked it even sooner, "there's no way to pressure the academy," given its thorough peer review.

With only a day to draw on the study for ammunition, Boehlert, Gilchrest and their allies introduced amendments substituting language on wetlands definition more in line with the academy report. But these were defeated easily, and by the end of last week the bill appeared to be on the verge of approval.

The debate now moves to the Senate, which may be more inclined to incorporate the academy's conclusions into its own bill, or even shelve consideration of the Clean Water bill temporarily rather than wade into the wetlands versus land owners debate. Administration officials have threatened a veto if the House version of the bill makes it to the president's desk.

Despite the furore in Congress, Lewis says the panel members themselves argued very little over the scientific issues, and that the report represents a "very deep consensus" among wetlands experts.

The irony is that Congress specifically asked the academy three years ago to help settle a question that had become politically intractable. But that was before conservative Republicans took charge of the House. Now, says a member of Gilchrest's staff, "they're not going to let anything as silly as science get in the way." **Tony Reichhardt**

Italian minister seeks more money for space

Munich. **Giorgio Salvini**, Italy's research minister, has promised to ask the government to increase the annual budget of the Italian Space Agency (ASI) from **IL850 billion (US\$503 million)** to more than **IL1200 billion** by transferring money from other government departments, in particular the ministry of industry.

Salvini told a meeting of the parliamentary committee for innovation and technology that he wants ASI to honour its international commitments, but also to have the same balance between national and international space projects as France and Germany. The space agencies of both these countries dedicate around half of

their budgets to national programmes.

Already **IL700 billion** of this year's Italian space budget has been allocated to international projects, mostly the European Space Agency (ESA) and others mounted bilaterally with the United States. As a result, little is left for national projects. ASI's director general, **Mario Calamia**, has been trying to renegotiate some international contracts.

But at the committee meeting Salvini reassured **Jean-Marie Luton**, director general of ESA, that despite its financial difficulties, Italy would "strive to maintain commitments it has already made to future ESA programmes". **Alison Abbott**

CSIRO attacked over review's 'omissions'

Sydney. The staff of Australia's largest research body, the Commonwealth Scientific and Industrial Research Organization (CSIRO), have launched an unprecedented attack on the board following the publication of an internal review last month that, they say, fails to address important issues.

A statement signed by more than a hundred CSIRO employees says there is a "widespread lack of confidence" in both the board and the committee — consisting primarily of board members — that produced the internal report.

Problems cited include the fact that the internal report failed to mention either communication problems within the organization or that scientific excellence was one of the organization's goals. The report is also said to have omitted any discussion of the need to restructure the board itself.

Roy Green, acting chief executive of CSIRO, has responded by promising that a "representative committee" will be set up to review internal communications and identify time-wasting procedures that could be abolished. But he has rejected the suggestion that the committee had not considered scientific excellence in its review.

A spokesman for CSIRO, which has an annual budget of A\$662 million (US\$490 million), said last week that the new committee would be similar to those often estab-

lished by the organization, but that no members have yet been named.

CSIRO's present problems stem partly from the adoption of a new management structure in 1986, under which the organization's basic operating units — known as divisions — were grouped into six 'institutes' of broadly similar research areas. Although this was meant to create synergy between researchers in related fields, the institute became directly involved in assessing the quality and effectiveness of the research carried out in the divisions.

Last year, a Senate committee report on the organization pointed out that the institutes generated "a vast amount of paperwork without clear and worthwhile objectives" (see *Nature* 372, 585; 1994). The committee also criticized the organization's board, chaired by Adrienne Clarke, professor of molecular biology at Melbourne University, for neglecting the employment conditions and morale of its staff, as well as the response of John Stocker, CSIRO's former chief executive, to the problems it had raised. Stocker left the organization at the end of his five-year term in March.

In response, CSIRO released in mid-April the conclusions of an internal review that broadly recommended abolishing the institutes and suggested a return to the old structure of divisions reporting direct to

head office (see *Nature* 374, 587; 1995).

At the time, Green denied that the review's recommendations were connected with the Senate report, and said that the institute structure had worked well, even if it was no longer appropriate. He also commented that any major changes would have to await the appointment of the new chief executive officer.

CSIRO staff are certainly anxious to see change. Peter O'Donohue, general secretary of the CSIRO Division of the Commonwealth Public Service Union, agrees that there is widespread support for the proposal that the institutes be abolished. But he emphasizes that it does not represent union policy, which is currently being thrashed out.

Mark Lawson

A modest boost for Australian science

Sydney. Australia's two major grant-giving bodies are the main beneficiaries of a slight increase in overall funding for science in the national budget announced last week.

Funding for research through the Australian Research Council is being increased by 9 per cent in real terms to A\$350.2 million (US\$260 million). Some of the extra money, however, has been earmarked for higher education. The National Health and Medical Research Council (NH&MRC) will have its allocation increased by 7 per cent to A\$139.2 million.

According to a 'Science and Technology Budget Statement' released the day after the main budget, the losers include rural research and development, which will fall by 6 per cent to \$130 million, and "other R&D agencies", whose funding will drop by 5 per cent to A\$232.2 million. Overall, however, the total federal spending on science next year will grow by 0.7 per cent to around A\$3.6 billion.

The statement contained some disappointments for researchers. In particular, although the NH&MRC's budget was increased — and despite some intense lobbying during the budget process — the new money is far from the extra A\$200 million a year which medical researchers believed they had been promised during the last federal election campaign.

Also, statistics presented in the budget statement show that Australian business spending on R&D is still much less than that of many other industrialized nations, falling as a percentage of gross national product below Italy and just above Ireland. One reason is that Australia's economy is still based largely on the mining industry, which tends to spend money on geological exploration rather than research.

M. L.

See also: pages 177 to 182.

Solar craft reaches for new heights

Boston. A team of scientists and engineers from AeroVironment, an engineering company based in Simi Valley, California, is about to begin testing a new version of the Pathfinder, a 400-pound (181 kg) aeroplane that derives all its energy from sunlight (right).



Test flights will begin in June at California's Edwards Air Force Base. AeroVironment is planning for high-altitude flights in late July, when it hopes to get the unmanned craft up to 65,000 feet (20,000 metres) and speeds of 100 miles an hour.

The original Pathfinder was built in the early 1980s, and reached a altitude of 8,500 metres in 1983. The aircraft was then put in storage while its engineers concentrated on improving propulsion technology.

Advances in lightweight silicon solar cell arrays — which will be laid on top of the plane's 800-square-foot

(74 metres-squared) wings — and in motors and propellers will make the latest version twice as efficient as its 1980s predecessor. With twice as much power reaching the propeller from the same amount of solar energy, the craft should be able to attain heights of 65,000 feet or more.

The project is being funded by the National Aeronautics and Space Administration (NASA) in the hope that solar-powered planes can eventually be used for environmental missions, to measure the ozone layer, sample the upper atmosphere and track storm systems.

Steve Nadis

Indian project will safeguard genetic value of plants

New Delhi. The Indian government has launched a major project to develop genetic 'profiles' of plants used for medicinal purposes as protection against the potential theft and patenting of useful genes by multinational drug companies.

The US\$5-million project is part of a programme on conservation of medicinal and aromatic plants funded by the Department of Biotechnology, under which a gene bank will be established in the Tropical Botanical Gardens and Research Institute (TBGRI) in Trivandrum in Kerala state.

The bank, which will maintain the DNA library, will also keep in a state of cryopreservation the tissue, pollen and seeds of plants identified after rapid screening as having medicinal value.

A trust named after Rajiv Gandhi, the late prime minister, has provided funds for the purchase of equipment, while three specialized microbiological institutes will assist TBGRI in carrying out the DNA profiling. The gene bank will be completed by 1998, and will contain up to 10,000 samples, with their genetic profiles.

The creation of the gene bank is a direct result of the biodiversity convention signed in Rio de Janeiro in 1992, which gives sovereign rights over plant genetic resources to the states in which they are found. "The DNA fingerprints will help us stake our claim if a foreign pharmaceutical company seeks a patent for a drug based on a plant species taken out of India," says Palpu Pushpangadan, director of TBGRI.

Pushpangadan points out, for example, that the widely patented neem products distributed in the United States are based on extracts from the neem (*margosa*) tree imported from India. Similarly, the world's entire supply of isabgol (*pysllium*), a traditional remedy for constipation, comes from the Indian subcontinent. But the government receives no royalties for isabgol products processed and sold by multinational pharmaceutical companies.

Pushpangadan says that although many medicinal herbs have already been taken out of the country, "India is still left with a rich genetic diversity that must be conserved and protected from theft". He says his gene bank could supply material for research by foreign companies, but on condition that India holds the patent on any resulting product.

Given the country's new patent regime, and the lack of its own resources for developing synthetic drugs, India's main strength lies in its medicinal plants, he says. For example, his institute is planning to file a patent on a drug from *Tricopos eylanicus*, the Indian equivalent of Chinese ginseng.

K.S. Jayaraman

Even excellence is no longer good enough for the NSF

Washington. A senior biophysicist is taking the National Science Foundation (NSF) to task over its grant selection process which, he says, often fails to fund proposals that are graded as 'excellent' in some fields, while finding money to support lower-grade proposals in others.

Oleg Jardetzky, director of the magnetic resonance laboratory at Stanford University in California, says that in the past year NSF's molecular biophysics programme has rated three proposals from his laboratory as 'excellent' or 'very good', only to reject them.

Having collected evidence that the funding situation in other programmes is less fiercely competitive, Jardetzky has written to Neal Lane, director of NSF, saying that he may offer testimony to Congress that NSF's funding process is "an arbitrary, and hence wasteful distribution of public funds".

NSF says that all its programmes are heavily oversubscribed, and the success rate in the biophysics programme last year, although low, was not abnormally so. But the agency, which spent more than \$2 billion last year on university research, says that it expects many more complaints like Jardetzky's. "As money gets tighter, this issue is going to become more important," says Anne Petersen, deputy director of the NSF.

According to Jardetzky, the biophysics programme was able to fund only 18 proposals last year out of 28 graded 'excellent' and 79 considered fundable — a success rate of 22 per cent. In a letter responding to his initial complaint, Mary Clutter, assistant director for biological sciences at NSF, said that 30–50 per cent of applications were considered fundable across her directorate, and that overall about a quarter of all applications received by the science agency were funded.

Jardetzky interprets this as meaning that more than half of suitable applications are in fact funded — and wants to know why only one-fifth of such applications succeed in biophysics. But Clutter says it is unfair to draw a comparison between broad estimates for the whole of the directorate and specific figures which only cover one year in one programme.

In some disciplines, she says, 80 per cent

of proposals are considered fundable. The overall success rate of 25 per cent varies widely between programmes, the variation depending on factors such as the extent to which NSF's funding responsibility overlaps with that of other agencies.

In many programmes, more proposals are graded 'excellent' than can be funded, NSF officials say. But these programmes are not necessarily those with the highest priority; they may simply be the ones whose study panels are most inclined to give 'excellent' grades.

Although the NSF tries to impose common assessment standards across all fields, equivalence between 'excellent' proposals in, say, electrical engineering and cell biology, may mean little.

That is why the NSF, as it has expanded over the years, has moved towards a system where it divides its money into separate programmes. The result, according to Jardetzky, is that the biophysics programme has become "a straitjacket", unable to meet the needs of scientists who depend on it.

Petersen says that some members of the NSF's governing body, the National Science Board, agree with Jardetzky that it would be better "just to take what comes in and fund the best work". But the size of the NSF today — as well as its accountability to Congress — makes a quick return to this utopian principle unlikely.

NSF also says that an independent 'committee of visitors' regularly inspects the handling of research proposals in each division. Two years ago, such a committee looked at the division of molecular and cellular biosciences, including the biophysics programme. The committee came to the conclusion that it was happy with the overall quality of review, but noted that "marginal budgets result in a great deal of excellent science being either underfunded or unfunded".

Furthermore, Jardetzky's threat to testify in Congress against NSF's funding practices may prove to be empty, as the science agency does not have any real enemies on Capitol Hill to take him up on it.

But he does promise to raise the matter with the National Academy panel, headed by former president Frank Press, on how the US government distributes its science funding (see *Nature* 372, 5; 1995). And whatever the Press panel makes of Jardetzky's complaint, it is still expected to address the growing problem faced by science funding agencies: how to maintain the peer review system intact when it is not possible to fund all the work that peers deem excellent.

Colin Macilwain

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

National Science Foundation

**Petersen: issue will
grow in importance.**

Asteroid hazards stir up defence debate

Boston. After last summer's fiery collision between Comet Shoemaker-Levy 9 and the planet Jupiter, concern about the potential impacts from large asteroids and comets is continuing to grow. Next week, an international conference of astronomers, planetary scientists and weapons specialists takes place at the Lawrence Livermore National Laboratory in California, to assess ways of defending the Earth from potentially devastating impacts.

The conference comes on the heels of a meeting in April, at which scientists and officials of the United Nations in New York discussed the hazards posed by near-Earth objects (NEOs) and ways to identify and track them. This itself was a follow-up to workshops sponsored by the US National Space and Aeronautics Administration (NASA) in 1991, which culminated in the January 1992 *Spaceguard Survey* report.

David Morrison, an astronomer at the NASA Ames Research Center in California, who chaired the 1991 sessions, told the UN meeting that a comprehensive search was needed for Earth-crossing asteroids and comets as outlined in the Spaceguard proposal. He claimed that a coordinated, international network of telescopes to carry out this survey would be "a good investment", considering that a major impact could kill millions of people.

"Right now, there are fewer than a dozen people searching for NEOs worldwide, fewer people than it takes to run a single McDonald's," says Morrison. That effort, he said, should be compared to the fact that the average annual risk of death from cosmic

bombardment is comparable to an individual's risks of dying in a plane crash.

Many claim that recent advances in charge-coupled device detectors and computers mean that a survey could be accomplished more quickly than the 25-year, \$300-million Spaceguard plan. Ted Bowell, an astronomer at the Lowell astronomer, told the UN meeting that an updated system might complete the job in about 10 years, and for about half the cost. "The whole thing could be done for the cost of a [cheap] space mission," he says.

John Remo, a physicist who organized the UN meeting, is working with the UN to ensure that the defensive aspects of such a programme, if approved, conform with international law over the use of military weapons in space.

A set of policy recommendations drawn up by Remo and approved by most conference participants has already won preliminary support from the UN Office of Outer Space Affairs. These call for the establishment of "adequate observational facilities" in both the northern and southern hemispheres, a mechanism for sharing information on potentially threatening NEOs, and broad international agreement before responding to such a threat.

He is planning to take a draft of the recommendations to next week's workshop at Livermore which is being sponsored by bodies ranging from NASA and the US Air Force Space Command to various Russian energy agencies.

According to Gregory Canavan, a physicist at Los Alamos who is one of the event's

chief coordinators, the conference is intended to complement the UN conference. "We're trying to integrate detection and mitigation [or defense] technologies to make sure that both approaches will be compatible," Canavan says. "In the past, people really haven't connected the two."

Canavan says the Livermore conference will not propose an actual defensive system. "We're hoping to define the problem in sufficient detail to identify the key technological components that will be needed, along



with their weaknesses, so that a few years from now we could put together a detailed design."

In the interim, he says, experiments could determine the strength of meteoritic materials, to "give us a sense of how hard you could push an object without breaking it up". 'Coupling' experiments, aimed at finding efficient ways of delivering energy to threatening NEOs, are also needed, he says.

Canavan says that an effective defensive solution will not be a single system, but will require a "grab bag" of techniques whose efficacy will depend on the size of the approaching asteroid or comet, and the time remaining before a projected collision.

Given time to prepare, one option might be to set up a huge 'sail' on the NEO to perturb its orbit, or to install solar cells that would power 'ion thrusters'. But Canavan says that, for a big object close at hand, nuclear explosions may be the only viable option.

But the idea of putting nuclear weapons in space remains highly controversial (as, indeed, does the whole idea of funding a major observation/defense programme, which some critics claim would be an inappropriate allocation of resources).

Such a measure would violate both the 1963 Limited Test Ban Treaty and the 1967 Outer Space Treaty. "The last thing we want to do is trigger another nuclear arms race on account of near-Earth objects," says Remo. "That would create a peril far greater than the NEOs themselves."

Steve Nadis

Nuclear storage decision postponed

London. The British government has postponed plans to build an underground nuclear waste repository close to the Sellafield nuclear fuel reprocessing plant in northern England "until at least the latter half of the next century".

Construction of the £2-billion (US\$3.2-billion) depository, designed to house low- and intermediate-level nuclear waste which remains radioactive for several thousand years, was expected to begin by 2005. But John Gummer, Secretary of State for the Environment, announced last week that building work on the structure, to be sited 650 metres below ground, will not begin before the year 2060.

"We are aware of the concern expressed by environmental groups and others," Gummer said in a written reply to a question in parliament which coincided with the government's announcement of its plans to sell off parts of the nuclear industry to the private sector. "Options available to future generations should not be foreclosed by irre-

versible action taken now," he said. A spokesman for the Department of the Environment (DoE) later added that the repository might be built even later than 2060, if the government chose to locate it at a site other than Sellafield.

Safe storage of nuclear waste is the responsibility of UK Nirex Limited, which is jointly owned by British Nuclear Fuels, Nuclear Electric, the UK Atomic Energy Authority and Scottish Nuclear. Attempts at a solution have been dogged by controversy and delays leading to mounting bills — and accumulating waste. Nirex has already spent two years and £50 million investigating 500 possible sites in the United Kingdom.

A white paper on radioactive waste management is expected later this summer. But a DoE spokesman said the decision to put off building the repository was released one month ahead of schedule to reassure potential investors in nuclear power that they would not be saddled with the job of disposing of nuclear waste.

Ehsan Masood

Germany to launch new gene programme ...

Munich. Germany is at last to become a full participant in international efforts to decipher the human genome, having been held back for years by social opposition to genetics research rooted in the nation's experience during the Nazi era.

Later this month, the research minister, Jürgen Rüttgers, is due to announce details of a new genetics research programme, costing DM50 million (US\$72 million) a year for eight years. The programme will give priority to human genetic diseases, and will be geared towards developing diagnostics and therapeutics.

Human genetics has not been completely ignored in Germany. But, unlike countries such as the United States, France and Britain, such research has been carried out by isolated groups without central organization or the strong support of government funding agencies. The new programme will now place German efforts on the same level as those of its partners.

According to Frank Laplace, who is responsible for the programme within the ministry of education and research (BMBF), one important goal will be to set up a resource centre to provide genetic probes for researchers and an extensive gene database. This will be based on the work of Hans Lehrach, a new director at the Max Planck Institute for Molecular Genetics in Berlin, who helped to develop the proposed programme.

Lehrach's group has been building a gene and clone database and an interactive service that provides scientists with probe filters for gene localization and clone identification. Clones, therefore, can be standardized between laboratories, which means "it is [a] much more efficient system for identifying large numbers of clones than sequencing," says Lehrach.

The programme will also provide competitive grant money whose extent and detailed priorities will be discussed at a meeting of the programme's international science advisory committee this week.

China aims to treble research funding

London. The Chinese government last week announced plans to treble investment in scientific research. But a rhetoric-weary science community may be in little rush to celebrate, as an apparent decision to increase funding on research and development from 0.5 per cent to 1.5 per cent of its gross domestic product was not accompanied by a detailed timetable.

The decision was announced at a meeting of China's ruling State Council, which announced that the target figure would be reached "by the year 2000". □

The prospect of a German genome programme has been widely welcomed, both nationally and internationally. Bertrand Jordan, for example, of the Centre d'Immunologie de Marseille Luminy in France, a member of the programme's international scientific advisory board, says he is pleased that active opposition to genetics research in Germany has declined.

Jordan also claims Germany's late start has some advantages. German researchers "can assess the situation, and are not tied to existing projects," he says. "This should allow them to move forward more quickly."

This opinion is not universally shared. Sir Walter Bodmer, director general of the Imperial Cancer Research Fund in the United Kingdom, and a former president of the Human Genome Organization (HUGO) in

Europe, says that starting late "means time lost with training and becoming involved". But he still welcomes Germany's commitment to human genome research. "It will add strength to what is being done in Europe", he says.

Daniel Cohen, director of the Centre D'Etude du Polymorphisme Humain (CEPH) in Paris, says that the new programme means that Germany has "finally stopped using history as a pretext not to do research in genetics", and that he expects German involvement to lead to improved sequencing and other technology.

"There have been no new developments in the last ten years," says Cohen. "The Germans are good at organization on a large scale, and they do very good research."

Toni Feder

... as France risks being left behind

London & Paris. One of the first decisions facing the government of France's new president, Jacques Chirac, due to be announced this week, will be its response to recent initiatives in the United States and the United Kingdom to start sequencing the entire human genome (see *Nature*, 375, 93; 1995).

France has become a major player in both gene mapping and gene sequencing, in particular, through the work of the Paris-based Centre d'Etudes du Polymorphisme Humain (CEPH) and the Génethon laboratory, and with a substantial input in funds from the French Muscular Dystrophy Association (AFM). But many observers feel that it, nonetheless, lacks a coherent national strategy, or even a consensus on how to move forward.

Fears that France could be left behind if the joint proposal by US and UK groups is adopted are reflected in a recent report commissioned by the AFM. This report claims that such a project involves not just scientific and medical interests, but also considerable financial and industrial stakes — pointing out, for example, that the rights to a gene associated with obesity were recently sold for \$20 million.

It warns that the launching of an American programme — either unilaterally or bilaterally with the Sanger Centre (see *Nature* 375, 93; 1995) — would place France in a "position of dependence" and risk giving US companies a competitive advantage. Such dependence, it suggests, could restrict French scientists from getting access to data.

The report also warns that tying into an international programme could potentially create conflict with France's own research priorities, while French research agencies

are concerned that the big money needed for 'megasequencing' would reduce their budgets for other priority topics.

One possibility, suggests the AFM report, is that France might launch an independent effort to sequence the entire genome in parallel with any US/UK effort. But given the costs involved, others see greater possibilities in international collaboration, and say that they are keen to contribute to the type of project being outlined by British and US research groups.

Whatever happens, everybody agrees that the government needs to decide, and soon, how much public money to make available. "The French should be active in this project," says Daniel Cohen, the head of CEPH, adding that "certainly there is a danger that France could be left behind."

Bernard Barateau, president of AFM, describes it as "astonishing" that the recent presidential campaign apparently ignored "one of the most massive scientific medical and industrial revolutions" of the end of the century. "It proves that the politicians aren't up to date, but it's not too late to catch up," he says, adding that France needs a strong domestic programme if it is to be a credible partner in any international effort.

But although Barateau's warning that France risks being left behind may appeal to the chauvinistic instincts of the new Gaullist government, the possibilities for action may be hindered by the confusion that now reigns in French genetics.

The previous government stripped the national genome agency of its powers and resources, while the money needed for AFM to repeat Génethon's mapping achievements is this time out of its reach.

David Dickson & Declan Butler

Berlin and global warming policy

SIR — Your leading article (*Nature* 374, 199–200; 1995) rightly emphasizes the importance of the global warming issue and the need for international negotiations to have “the whole world on side” both scientifically and so far as policy is concerned. We would suggest that one of the main reasons why global warming is being taken seriously by governments is due to the work of the Intergovernmental Panel on Climate Change (IPCC) which has provided effective, unbiased and widely available assessments of the science of climate change, its potential impact and possible adaptation and mitigation strategies. Your leading article argues, however, that the IPCC should be replaced in its function of giving scientific and technical advice by a body based on the model of the UN Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). This view seems to be based on superficial knowledge of the IPCC reports, misunderstanding of what the Climate Convention (the FCCC) requires and ignorance of the *modus operandi* of the IPCC.

One reason that you cite is that the IPCC “reflects the weaknesses and vested interests of its sponsors” (what these are is not explained). We can categorically state that the IPCC determines its own programme of work and publishes its assessments with no interference from its sponsors, the World Meteorological Organization and the UN Environmental Programme.

A further reason you mention concerns the mode of working of IPCC. In fact IPCC carries out its work largely in the way you propose. The IPCC plenary (composed of government representatives) determines the scope of the reports, full-time staff is available in Geneva (for central coordination) and at the three IPCC working group secretariats. It has a large army of ‘consultants’ — leading members of the research community who donate their time and who are primarily responsible for drafting and reviewing the technical reports. Workshops and seminars are organized on important issues. The very wide participation by the scientific community in the preparation of the reports is one of the strengths of the IPCC process compared to one (as in UNSCEAR) where the report’s preparation is in the hands of comparatively few. Scientists from developing countries are also included.

A further point you make is that the IPCC assessments are not “comprehensive or welcoming of contrary opinions”. No convincing evidence is presented to support this. In fact the scope of the IPCC assessments is very broad and comprehensive. The reports cover all aspects of

climate change including perturbations to the physical system, impact on both ecological and socio-economic systems, technological options for adaptation to climate change and for the mitigation of emissions, geo-engineering options (which are not always the “crackpot ideas” you suggest but which deserve serious analysis) and methodologies for use by countries in developing their inventories of greenhouse gas emissions and their potential impact/response options. The Second Assessment Report currently being prepared by the three working groups will contain nearly 60 chapters.

Regarding contrary opinions, the review process undertaken by IPCC specifically addresses conflicting scientific opinion and ensures that these are properly exposed and represented in the statements of uncertainty. A particular example quoted in the leading article concerns the role of water vapour which, because of its importance as a greenhouse gas, is crucial in any assessment of global warming. It is stated that water vapour (especially upper level water vapour and ice) is neglected in climate models. This is simply untrue. As is evidenced in all of the IPCC reports, models that have been available during the past five years include comprehensive descriptions of all phases of water and of their role in cloud formation. Further, because of some significant controversy over the magnitude and the sign of water vapour feedback (associated with the ‘air conditioner’ idea mentioned in your leading article) the IPCC 1992 report discussed the role of upper tropospheric water vapour in some detail. The considerable uncertainty that still exists about the magnitude of the feedback is reflected in the uncertainty range given in the IPCC reports for the magnitude of the expected global warming.

The IPCC is a relatively young institution and a great deal of learning has taken place during the first five years of operation. We are open to further improvements. But the structure and procedures of the IPCC have allowed the scientific and technical community to be responsive to governments while pursuing what it believes to be scientifically important.

The IPCC is charged with providing scientific and technical information to the public and to policy-makers but not to propose action programmes. That is for the political process. For instance, whether to act soon or not is a political decision that must be made in the light of careful analysis of the inertia of both the

global environmental system and the socio-economic system, as well as an evaluation of the prospects for developing and applying technologies that permit continued economic growth with lower rates of greenhouse gas emissions. Such analyses are provided in the IPCC assessments.

The IPCC process has two main strengths. First, in its assessments it has succeeded in involving a very large proportion of the active scientists (many of them world leaders) working in the field over a wide range of countries. IPCC reports have therefore achieved a large degree of ownership by the worldwide scientific community. In a field as complex and controversial as global warming, assessments by smaller groups of scientists would not carry the same degree of acceptance. Second, because IPCC is intergovernmental and involves government scientists in its approval process, it has also been recognized as their prime source of information. This is absolutely essential. To move the policy debate forward requires the broad acceptance by governments of the underlying science. Any body chosen to advise the UN Framework Convention on Climate Change must ensure that these two strengths are maintained.

Bert Bolin

(Chairman)

John Houghton

(Co-chairman of Working Group I)
Intergovernmental Panel
on Climate Change,
Meteorological Office,
Hadley Climate Centre,
London Road,
Bracknell RG12 2SY, UK

Sounds right

SIR — The scholarly analysis of Katsikis and Leben (*Nature* 374, 670; 1995) seems to have turned the tide Funder’s way on apoptosis.

We write to point out that he unnecessarily weakened his own case in his opening letter (*Nature* 371, 98; 1994) by implying that Italians would not pronounce the second p when they say apoptosis. Surrounded as we are by Italians working on apoptosis, we can assure him they do.

The examples Funder cited were the Italian equivalents for pterodactyl and helicopter, in which he claimed the p is neither written or sounded. He was right about helicopter (*elicottero*), but wrong about pterodactyl (*pterodactilo*). The p is written and, like anything written in Italian, it is sounded.

John Ross

David Burr

Istituto di Neurofisiologia del CNR,
Via S. Zeno 56,
Pisa 56100, Italia

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

Australian science looks up and ahead

A brief visit to Australia confirms the impression that there is much excellent Australian science; the surprise is that the research community has embraced the goal of wealth creation with enthusiasm and some success.

AUSTRALIA is comparable with the Netherlands both in population and gross national product (GNP), but with three striking differences. First, Australia is a continent in its own right and not a narrow strip of land half reclaimed from the sea. Second, Australia is far from most other places and therefore geographically prone to the feeling of neglect engendered by isolation. Third, unlike the Netherlands, Australia is now growing quickly in size, as much from immigration from the north as from indigenous growth, and becoming markedly more cosmopolitan in the process. To most outsiders, it is also important that the Sun is in the north at midday.

What follows is an impression of Australian science formed on a brief visit to Australia early in April, mostly to places familiar from previous visits. The general conclusion is that Australian science is in good shape. That is not altogether surprising, given the attention lavished on the question of how to improve the condition of research over many decades. The way in which pessimism has now so often been replaced by optimism suggests that Australia may at last have found the right arrangements for the management of publicly supported research. But none of that implies that the ideal has been attained, or that many people in Australia harbour that illusion.

The reasons why the management of science is given more public attention in Australia than almost anywhere else are interesting, important and inescapable. Two stand out. First, Australia is an intensely political place. The federal (or 'Commonwealth') government at Canberra holds the centre stage, but each of the state governments has a comparable structure (with a prime minister and an elected parliament), mostly organized along British lines. (At the end of March, newspapers in New South Wales were paying as much attention to the prospect that the newly elected state parliament might be 'hung' as they would have done to a similar outcome of a federal election.) Then there is local government as well.

Democratic inquisition

But the Australian system differs from the British in that there are fewer restrictions on the freedom of members of a parliament to enquire into the uses made of public funds, even when they are spent by

supposedly autonomous boards. On the contrary it is standard practice for such organizations to be repeatedly reappraised, often at the whim of a new minister or at the urging of a parliamentary committee. The chief victim of this process has been the Commonwealth Scientific and Industrial Research Organization (CSIRO), once virtually the sole

have won prestige fairly. Elitism evokes instinctive mistrust.

Mr Jones's legacy

The second explanation of the attention now paid to research in Australia is the growing recognition that science has an important part to play in Australia's future. There has been a huge change in



research agency in Australia, but still conspicuous because its annual budget is substantial and because its work touches much of Australian life.

There is no doubt much in the opinion of many Australians that repeated external enquiries into the working of even specialized public institutions are a symptom of democracy in action. Australians are proud both of their outspokenness and their willingness to discard outmoded institutions, however much hallowed by tradition. (The British monarchy may, in that spirit, soon be discarded.) One unwelcome consequence is the habit of correctitude from which public institutions cannot escape, and bureaucracy that feeds on that. Another is the suspicion that there must be something wrong with outstanding institutions even when they

the past decade, a substantial change in merely five. At least part of the credit for that must go to Mr Barrie Jones, Australia's first minister of science (in 1983), who appears to have owed his appointment to a polemical book, *Sleepers Awake!*, which urged Australians (the "sleepers") to take their proper place among the quickly growing nations of the Pacific Rim by investing in advanced technology.

During his too-short spell as science minister, Jones did much to systematize the collection and publication of data about the Australian research system; that survives his tenure. But his more important achievement was to be publicly recognized as an enthusiast for research projects, the Australia Telescope (see below) among others. Jones would repeat-

edly turn up at inaugural ceremonies to which most ministers would have sent a civil servant. Inevitably, he instituted a reorganization of CSIRO.

His failings as a minister appear to have been wayward administration and his difficulty in explaining what he was driving at in language his fellow-ministers could understand. As one supporter said last month, "Barrie said all the rights things, but nobody could understand him". Jones remains an elected member of the federal parliament and is currently the president of the Australian Labor party.

One sign that Jones's message has partly sunk home is the prevailing Australian habit of referring to the country's lowly position in the league table of its members' research spending maintained by the Organization for Economic Cooperation and Development (OECD), in which Australia has long come fourth from the bottom (before the accession of Mexico last year), above New Zealand, Ireland and Spain, with research and development expenditure amounting to just under 1.30 per cent of GDP. (The comparison with the Netherlands, where the same ratio works out at 2.16 per cent, breaks down at this point; Japan is a long way ahead at 2.90 per cent.)

Industrial research

Another good sign is that industrial research spending has indeed been growing quickly, more quickly than in any other OECD country, and now comfortably exceeds A\$2 billion a year. Industrial finance for research has grown to the point at which it almost equals the government's spending. These comparisons may also be unflattering to Australia in that Australian salaries are low by international standards. (One researcher complained that when he visits London now, he cannot take out an ex-graduate student to eat "even at an Indian restaurant" without spending more than it would cost in Australia to entertain his family).

Yet another sign of reviving industrial interest in research is more down to earth. Last month, Australian researchers were marvelling that a senior colleague at the renowned Radiophysics Division of CSIRO had resigned to take an industrial post at a salary three or four times his government pay.

Some of these trends have been impelled by the 150 per cent tax credit for industrial spending on research (the deduction of that proportion of research spending from corporate revenues before taxes are calculated). This measure, which goes back to Jones's days, is estimated to cost roughly A\$400 million a year, but has been declared a permanent feature of Australia's system of corporate taxation.

More direct, and perhaps less costly, ways of encouraging industrial interest in

research seem also to have had an important influence.

Collaborative research

One of these, for example, is a programme for founding Collaborative Research Centres (CRCs) involving groups at academic institutions with others from industrial environments. Since the beginning of the scheme in 1990, more than 61 centres have been founded. The government pays half the cost of the collaborations, which mostly goes to the academic members of the partnerships. Independent government agencies can join the partnerships; roughly half of those so far established include divisions of CSIRO. Costs to the government can be quite modest, perhaps only a few hundred thousand dollars (but the Biopharmaceutical Research Centre based on the University of New South Wales costs the federal government nearly A\$2 million a year).

Part of the distinctiveness of this scheme lies in the way in which the centres are managed. They are constituted as if they were commercial joint ventures, and managed by a director and representatives of the partners acting as a kind of board of directors. Among other things, they provide research programmes with timetables and agree in advance on what decisions should be made and by when in the course of particular projects, which may within the context of a broader framework be short-term affairs.

Many actual projects are severely practical. For example, the Plant Science Centre (based in Canberra) involving the Australian National University, the Plant Sciences Centre of CSIRO and the commercial company Group Limagrain Pacific, has an overall interest in the genetic manipulation of wheat, but has found time to produce a means of using DNA polymorphism (made visible on gels) to distinguish between 18 varieties of cannabis. The objective is not the better to meet the needs of illicit drugs users, but to provide a means by which law enforcement authorities can tell when cannabis is being grown legitimately (it is widely used in making special paper, for example).

The same centre is also the custodian of Australia's interest in the international programme to sequence the *Arabidopsis* genome and, nearer the country's cultural and economic interests, the genetic improvement of malting barley and the project to develop international conventions based on DNA polymorphism to tell when plant varieties are sufficiently distinct to justify registration (and then to tell different varieties apart objectively).

Some academic institutions have joined CRCs as if membership were a collectors' item. Thus the University of New South Wales (in Sydney) belongs to no

fewer than 11 centres, dealing with topics as different as 'cardiac technology', the manipulation of thin sheets of metals or of composites in the construction of the skins of aerospace structures and photonics (which means fibre optics). The University of Melbourne belongs to 10. One centre at the University of New South Wales, for Waste Management and Pollution Control, whose non-academic members are mostly public utilities in New South Wales, has already been incorporated as a limited company.

There is no denying officials' claim that the CRC scheme has been a great success. An important part of the explanation is that the arrangements bring comfort to industrial and academic partners alike. The former are delighted with the way the research programmes are managed on industrial lines, the latter with the flexibility provided by the freedom to change particular objectives within the overall terms of reference.

Academics have also been won over by the discovery that research frameworks settled only after negotiation with industrial partners turn out to be more interesting than originally appeared possible. Thus the Industrial Plant Biopolymers CRC based at the University of Melbourne, whose objective is to find ways of making commercially important polymers such as gums from plant cells in culture, may yet provide the first compelling explanation of why these materials have such remarkable elastic and adhesive properties.

The success of this scheme owes a great deal to the technical modesty of the objectives by which the centres have been specified. That is not to say that the research is dull; participants are ready to proclaim the opposite. And the plan (at the Biopharmaceutical CRC at Sydney) to mass-produce monoclonal antibodies for the treatment of leukaemia is hardly unadventurous. Rather, the objectives chosen have been attainable, thus allowing the partners to taste achievement early. Similarly named cooperative centres in Britain and the United States have often lacked that quality.

The fly in this ointment is that the future of the newly created centres remains unclear. The government has been uncharacteristically silent on the subject. Yet with the first allocations of funds made in 1991 for periods of five or seven years, the government must say before long whether it will continue with its support, both for at least some of the CRCs and for potentially new creations. It may also discover cause to go back on its original decision that the CRCs set up in 1994 will be the last.

The best strategy would be that the government should wait to see which are the CRCs whose industrial partners wish to abandon ship, to let them lapse, but

then to devise a realistic plan for telling how and when the remainder could become free-standing entities. Although CSIRO's presence in about 50 of the existing partnerships (at a total cost over the planned lifetime of about A\$350 million) may be taken as a substitute for continuing government support, it is mostly in kind, not cash, and in any case is only half as much as Canberra spends on the CRCs.

On the principle that the best investments are in success, it would be a great misfortune if the government were now to turn its back on CRCs, existing or still gleams in people's eyes. Whatever happens to this brave experiment in Australia, there is every likelihood that it will be widely copied elsewhere. Or it should be.

CSIRO

Another, and the most important, influence nudging Australian research into what the British would call 'wealth creation' is Australia's largest single research organization. Times have changed, at least since CSIRO enjoyed a virtual monopoly of research in Australia say a quarter of a century ago. (Many universities were naturally already active in research, but were often dependent on CSIRO or some other government department for all but routine research expenses.) CSIRO's researchers were then organized into laboratories called 'divisions', too many to be managed coherently. One consequence was that some divisions, notably the Radiophysics Division at Epping, outside Sydney, became too powerful for the good of the organization as a whole. And others (including some of the same) grew complacent.

Something of the old structure remains; the divisions continue. But since a bout of introspection a decade ago, divisions with similar interests have been grouped together into 'institutes' comprising between four and eight divisions each. Until, it seems, just now; last month, CSIRO announced the results of an internal management review whose chief recommendation is that the institute structure should be abolished to remove a whole layer of management. A final decision will await the recruitment of a new chief executive of CSIRO.

The organization's old and avuncularly exercised role as the chief provider of academic research funds too big to be accommodated by university budgets has now been taken over by the Australia Research Council (ARC), which has almost ten times as much to spend each year (about A\$300 million) than its predecessor, the Commonwealth Research Grants Council. But CSIRO is still responsible for managing as national (and international) facilities instruments

An exception that proves the rule

IN the land of Ms Germaine Greer, which is not much admired by feminists, the most influential working-scientist is a woman. Dr Adrienne Clarke is professor of botany and head of the School of Botany at the University of Melbourne. She is one of a handful of Australian scientists to have been the recipient of funds for a Special Research Centre from the Australian Research Council, which is worth just under A\$1 million a year.



Clarke: chairman twice over.

She has been chairman of CSIRO's board for just under three years, is a non-executive director of two start-up biotechnology companies, of Alcoa Australia and of a large international insurance company based in Australia. As well as that, she belongs to two Collaborative Research Centres in Melbourne, acting as chairman of that for Industrial Plant Biopolymers, as well as being a member of the absorbing Commission on the Constitutional Centenary which is due to recommend options for constitutional change in good time for 2002. (Arrangements for the Republic of Australia are in the front of many members' minds.)

How does one person manage all that? She says she has one unchangeable rule: never miss a pre-arranged lab meeting

about research. But she saves time by being quickly decisive, concealing the assertiveness that might suggest by a show of feminine diffidence. Tycoons are all alike, whatever their sex. But Clarke appears to have won the respect of her colleagues everywhere by being right most of the time, and sympathetic to the difficulties decisions always cause.

Her special centre (not to be confused with a collaborative centre, which is not nearly as grand) deals with the molecular basis of self-incompatibility in plants that will not allow their ova to be fertilized by their own pollen (thus undermining much of the advantage of sexual over vegetative propagation). Clarke and her lab-ful of postdocs are looking now for further identification of the RNase gene usually found in the region of the genome coding for the molecular constituents of the stylus. Quite why (and how) the presence of a presumably specific RNase in the stylus of a flower should determine compatibility is an open question, but has provoked a hunt to characterize the constituents of the stylus.

The postdocs are an interesting tale in themselves. In the roster given in the centre's annual report for 1993, nine out of 13 postdocs were from overseas, mostly from Germany, Switzerland and Japan. All five visiting scientists listed were from either the United States or Japan. There could hardly be more vivid proof that the world knows about the Melbourne compatibility lab. □

such as the Australia Telescope (see below), a newly completed linear array of radiotelescopes that is the only radio-interferometer in the Southern Hemisphere whose performance is comparable with that of the Very Large Array (VLA) in the United States. It also has Australia's ocean-going research vessel on its books.

Meanwhile, CSIRO has also been constrained by the federal government's requirement that it must raise 30 per cent of its budget from sources other than the parliamentary subvention, as well as by the more pragmatic decision that the government subvention will remain roughly constant from year to year (at A\$460 million a year in round numbers). By a blend of user fees (which some academic researchers say are too high for them to be able to afford) and research contracts (often with other government departments) industrial income is rising.

The result is that CSIRO has become an entrepreneurial organization willing to put a commercial price on the practical knowledge and insight at its disposal,

but whose interest in basic research, represented by the research interests of those who work for it, are as strong as ever, if more tightly, even carefully, limited. For practical purposes, CSIRO has been operating in this way for only three or four years. There is every reason why it should be given a chance to prove that its new marching orders will accomplish what the government intended by them.

New policies and also the succession of the generations have brought a profound cultural change. As recently as the 1970s, the organization's culture was a product of a least three distinct influences. There was CSIRO's commitment to traditional industries, notably in agriculture and mining. Sheep-farmers, for example, paid a levy on the wool they sold to the Wool Board, in return for which CSIRO did remarkable work on improving the genetics of the Merino sheep. The culture of a high-level agricultural extension service was almost palpable.

Another striking influence was that of the handful of British people who had

been radar practitioners in the Pacific during the Second World War, who decided to settle in Australia and who quickly turned the Radiophysics Division into a world-class laboratory for the design of sophisticated electronic equipments (which it is still). One was W. E. ('Taffy') Bowen, who became the 'Chief' (or director) of the Radiophysics Division after a few years. Another was Paul E. Wilde, who succeeded Bowen at Sydney and afterwards became the Chief Executive of CSIRO. Excellent as they were as scientists, they exuded the belief that that CSIRO was just another officers' mess, a friendly kind of club in which rank could occasionally become very important. But fierce loyalty to CSIRO was a hallmark of their manner.

Then there is what can only be called the modern influence, the culture of those trained in research only in the past 30 years, increasingly often themselves Australian. In the Australian style, these people are anxious to get things done, and quickly. They are also keenly attuned to what is happening in science elsewhere in the world. They are used to making deals to get what they want; wealth creation suits them, which is not to say that they are indifferent to the interest and the importance of basic research. A few years ago, one of the moderns, faced with a choice between commercial opportunities outside and uncertain prospects of promotion within, blurted out dramatically, "But I love CSIRO; I'd do anything for it!"

Inevitably, the modern influence is gaining strength. At the Melbourne half of the Division of Biomolecular Engineering (the other half of the division is in Sydney), for example, the declared objective is to "target the pharmaceutical industry" by seeking novel kinds of drugs. The protein structure group at Melbourne is that which, two years ago, decorated the cover of *Nature* with a model of the structure of influenza neuraminidase conjugated with antibody; the search, now is for a simpler chemical capable of binding to the active site of the neuraminidase molecule and there functioning as a prophylactic drug. The project is supported financially by Glaxo.

Further cultural change has no doubt been impelled by the changed composition of the Board of CSIRO, the organization's governing council. The old 20-plus board has been slimmed down to nine, with no fewer than six members from industry, business and administration. (The new chief executive when recruited will be a member *ex officio*.) This arrangement naturally makes for swifter and cleaner decisions. But conversations with three board members on a journey to Australia's telescopes suggests that they are anxious not to "harm the science", as one of them put it.

Government

In the Australian manner, the government has repeatedly reorganized its arrangements for the administration of public science. At the outset of the present string of Labor governments, Barrie Jones's ministry was free-standing, but that was soon absorbed into a larger ministry with responsibility for industry and commerce (making Jones into an unwilling second fiddle outside the cabinet proper).

Now, science is part of a portmanteau ministry with responsibility for industry, science and technology. The minister, Senator Peter Cook (a member of the federal parliament's second chamber), is a member of the cabinet. Although in post for only a year, he is well regarded. As always, his value to science will increase with the passage of time.

The most important (and expensive) part of Cook's brief is CSIRO, but the ministry is also responsible for what appears to be an effective means of coordinating the interests of other ministries in research. The key figure is Dr John Bell, deputy-secretary at the ministry and its chief scientific adviser.

But other ministries, of which the chief is the Department of Employment, Education and Training (known as 'Deet'), also have an important influence. Quite apart from its direct influence on universities through their budgets, Deet is also responsible, for example, for the Australian Research Council, which is a source of research grants to academic scientists. One of the most encouraging developments in the past few years has been the strengthening of this mechanism (see also page 172).

Biomedical research projects are most directly supported by the National Health and Medical Research Council, but on a disappointingly small scale (A\$130 million a year). It is true, of course, that this is a field in which endowment and charitable funds as well as hospital incomes make important contributions to research, while many groups with important interests in biomedical research are established as Collaborative Research Centres. Yet the sense that biomedical research just now is pushing at an open door is as vivid in Australia as it is elsewhere.

Australia's outwardly small defence research activities probably deserve more attention than they are usually given. Spending runs at just over A\$200 million a year, but is impressively diverse. One recent achievement has been the development of a long-range radar antenna capable of monitoring events on the southern edge of the Asian land-mass to the north (and well over the horizon), but the defence department has a distinctive record in the development of novel materials for ship construction and of materials such as cadmium mercury telluride as infrared sensors.

But in research, Australia is richest in its provision of structured introspection about its activities and motives. Since 1978, there has been an organization called the Australian Science and Technology Council (ASTEC) which exists to provide the government (and the world at large) with independent advice on technical issues. The chairman is traditionally an academic of some distinction. The council's work programme, either provoked by government requests or ASTEC's own judgement of what questions are worth asking, have proved as catholic as anybody could expect.

There is also a Prime Minister's Science and Engineering Council whose membership includes interested cabinet members (including the treasurer), intended to give point to important issues in science and technology. That in turn is mirrored by a committee composed largely of officials and called the Coordination Committee on Science and Technology.

But the reach of democracy's arm is long. The Australian government is well used to setting up *ad hoc* committees, usually called task forces, to deal with whatever awkward questions may arise. The results may be almost bewilderingly decisive. Brooding in 1991 about the reasons why Australia was backward in the commercial exploitation of research, for example, the government responded to two task-force reports by setting up an organization called the Australian Technology Group, modelled on the British Technology Group, whose function is to back patentable Australian innovations with development funds and then further to exploit them. The group is only in its second year of existence, but talks enthusiastically if secretively of the inventions it has tucked away up its sleeves.

What all this appears to betoken is that the Australian government is genuinely persuaded that Australia's future lies with science and technology — with the usual disclaimer that there is only so much that it can afford to spend at any time. As everywhere, a few demonstrations that research can generate prosperity would help to stiffen its resolve, but the present mood is a far cry from the sleepiness of the 1970s.

Yet the government and its agencies remain concerned that the proportion of Australian spending on research that comes directly from Australian industry's own pocket is small by international standards (at less than 0.50 per cent of GDP). But there are also good things to say about this pattern. Only a small proportion of industrial spending on research is provided by the federal government, while the scale of industrial spending is at least growing more quickly than the rest of the research economy. The cry, that industry does too little, is familiar else-

where, and is unlikely to be stilled until a few of the companies started in the past few years have made their mark on the international markets. But the government may be over-impatient.

Universities

Curiously, another important instrument of innovation in Australia, the higher education sector, appears to have survived in reasonably good shape the trauma caused five years ago by the reorganization of universities and other institutions decreed by Mr John Dawkins, then the minister at Deet. He began a process widely called 'Dawkin-sization' (see *Nature* 343, 203–206; 1990).

The declared objective was to broaden the base of Australian higher education, chiefly by arranging that institutions of higher education not designated as universities would share the privileges and distinction of the older institutions. Australia's established universities saw the intended reform as an attempt to cut them down to size; that calculation was probably correct. But a further objective was to reduce the cost of the administration of higher education by amalgamating small establishments with larger units, so reducing the cost of the academic overhead.

In democratic Australia, Dawkins's methods left much to be desired. Mergers between universities and, say, nursing or teacher education colleges were decreed from above and universities were left with the impression that their routine funds would be sharply reduced if they failed to take the initiative in carrying them through. At the same time, various colleges throughout Australia were designated as the cores of fully fledged universities, and given a remit to grow without noticeable restraint. There was vociferous opposition from long-established universities at Sydney and Melbourne.

Events have shown that many of the shotgun marriages decreed by Dawkins have turned out well. The University of Sydney's nursing school, for example, is now an integrated part of the university, and has not made academics previously in post believe that their careers have been demeaned.

Unsurprisingly, it has also turned out that the technical colleges designated as universities between 1988 and 1990 have indeed contributed to the education of the young, as was intended all along, but that they are not yet as successful in the competition for research funds from conventional sources as are the better-established universities. Industrial contracts are of course a different matter. Much the same pattern has emerged in Britain, where the previous polytechnics were designated as universities while the Dawkins reorganization was under way

in Australia.

That is one sign of what seems to have been the outcome of the upheaval five years ago; Dawkins won the battle but lost the war. The amalgamations were indeed forced through, but the established universities were stronger as a consequence. The paradoxical outcome of what seemed an egalitarian exercise has proved to be the opposite; the high achievement of the highly achieving universities is more than ever apparent.

So too is the palpable sense that Australian scholarship has finally conquered the sense of isolation with which it has grappled throughout the century. Not simply is travel now recognized to be a necessity (often made painful by jet-lag) and Internet's e-mail a fact of life. It is much more important that the size of the academic community is now great enough to ensure that it is confidently self-stimulating.

There is only one sad coda to this tale of how the university turmoil of the late 1980s turned out. Mr Alan Bond, best known for having bankrolled a successful Australian challenge for the Americas' Cup (for ocean-racing crews), but then an apparently successful tycoon, founded Australia's only private university, called Bond University, on Australia's Pacific coast.

The intention was primarily to provide young people from overseas, Japan perhaps, with a university education in a tropical climate within striking distance of the Great Barrier Reef. Professor Harry Messel, previously head of the School of Physics at the University of Sydney, became chancellor of Bond University before Mr Bond's bankruptcy two years ago. By all accounts, Messel is undaunted by this development. He has become "Executive Chancellor" of the university and plans to make a go of it. That is another feature of Australian academic life to watch in the years ahead.

Research

There is no way of singling out the distinctive elements of the pattern of research in Australia without the risk of seeming invidious or arbitrary, generalizing from too little recent observation. But some things are not disputed.

First, Australia has and deserves a high reputation in astronomy. The origins of this state of affairs are several. Until the exploitation of the Andes in the past two decades as a location for powerful telescopes, Australia was a natural location for the observation of the Southern Hemisphere, while Australia's practical interest in radar during the Second World War translated naturally into the development of radioastronomy.

CSIRO's powerful Radiophysics Division has had a hand in almost everything, beginning in the 1950s with the instru-

ment called the Mills Cross (after its designer) at the University of Sydney site west of the city. The same partnership continues with the radiotelescopes (operated as national facilities) at Parkes and at Narrabri, where the interferometer called the Australia Telescope is located.

That is unlikely to be the end of the matter. The Mount Stromlo Observatory near Canberra and the Siding Spring Observatory (in the north), both optical observatories, have been inherited by the Australia National University, which held an international meeting at the end of March to plan an international observatory on the Antarctic ice-cap.

The advantages of the location for astronomy (as distinct from astronomers) are plain: perpetual night for much of the year, the good seeing that goes with a ground temperature of -80°C and the prospect of being able to make observations in regions of the infrared not previously accessible through the atmosphere. The people are quite serious, and are even talking of carrying a 4-metre telescope to a height of 7 km above the ice (and 10 km above sea-level) by means of a tethered balloon. The next (and first practical step) in to install a geophysical sounding-station at the South Pole.

Second, Australia stands out for contributions to the science of the Earth. How could it be otherwise on the heavily weathered surface of an ancient continent, largely stable tectonically but otherwise well supplied with ore-bearing rocks from much earlier times in the Earth's history? But Australian Earth science has long since ceased to be driven by the needs of the mineral companies (interesting though they are).

Kurt Lambeck, until recently head of the School of Earth Sciences at the Australian National University (and still active there), is among other things distinguished for his refinements of the understanding of the rotation of the Earth. But the same laboratory is also the source of the ion microprobe instrument called SHRIMP, which makes it possible to determine the composition of single grains in specimens of crystalline rock, notably that of the zircons taken to be relics of early events in the formation of the Earth.

But a member of the same department, the late A. E. ('Ted') Ringwood (really a high-pressure physicist) was as much concerned with the messier business of using synthetic rock (called Synroc, a ceramic titanate) as a safe repository for radioactive waste.

The Chinese government is said to be interested in the process, which has been developed to the point at which 10 kg batches of Synroc, doped with simulants of radioactive waste, are synthesized from dry starting materials with the intention of finding out which recipes

serve best in different circumstances.

Inevitably, Australian interests have broadened outwards from this base. Surrounded as Australia is on all sides by water, oceanography and marine biology have flourished. The environment of the Great Barrier Reef remains a challenge, as does its origin. More recently, El Niño (previously a Western Pacific phenomenon) has turned out to have Australasian consequences. And the Antarctic is not all that far away.

The low density of human settlements, and the consequently small local contamination of the atmosphere, also makes the continent a convenient laboratory for the study of the changing composition of the atmosphere. Beach-going Australia has developed a keen interest in the stratospheric concentration of ozone.

The life sciences are also flourishing, and on several fronts. One traditional feature of Australian medicine is that it has always been more occupied with science than people have expected. It is no surprise that the hospitals of Victoria, of which Melbourne is the capital, were among the first in the world to take up the practice of *in vitro* fertilization and that the state's politicians followed quickly with legislation.

Nobody is now surprised, of course, that the life sciences have become molecular, but the University of Adelaide has gone further than most by giving its school of agriculture molecular foundations. For example, the Waite campus of the university has long since absorbed what was previously the Waite Agricultural Research Institute, using it for agricultural research and teaching (and the Waite family house as a faculty club).

The ambition is to teach a generation of students who will be able to practise agriculture with their eyes open to the modern world, while manipulating standard plant stocks genetically so as to improve their genetic potential. But this is not purblind molecular science. People are as much concerned with soil, climate and water supply as with the arrangement of nucleotides.

People matter in circumstances such as these. The life sciences at the University of Adelaide (which is not so much a city as a central suburb) owe much to Professor Harold Woolhouse, the immigrant plant geneticist from the United Kingdom who is credited with having carried through radical reform of the faculty in just over a decade's spell at Adelaide, and Professor Bob Symons, also a doer if in a less formal style, whose legend is that frustration at failing to buy radioactive nucleotides locally led him to set up a company to manufacture them which has become one of South Australia's industrial successes. One has the impression that these people travel as often to Europe as to Sydney, which is culturally as far away. □

Radio-telescopes and kangaroos galore

Narrabri. At dusk, the kangaroos come out onto the telescope campus here. It is a grand sight. They lope off in groups at great speed, their forelimbs almost vestigial compared with their powerful propellant limbs. Nobody has counted them accurately, but people say that "there must be several hundreds". The telescope managers say that the kangaroos do not interfere with the instruments, so that they need not be chased off. Neighbouring farmers, by contrast, find the kangaroos intolerable. The Australia Telescope (AT) has become their sanctuary.

The telescope is Australia's great pride. It is laid out on several kilometres of 6-metre gauge railway track, along which can be trundled independently steerable dish antennae configured for high-frequency reception. Two criteria matter for an interferometer like this: the pointing accuracy of the individual dishes, and the design of the electronics by means of which the separate signals are compared and synthesized into a map of a distant radiogalaxy.

The managers believe they have the design just right. For what it is worth, the AT is the only instrument of its kind in the Southern Hemisphere with comparable aperture and frequency range. The natural comparison is with the Very Large Array in the United States. (But the AT is for the time being a linear array; off-axis dishes may be added at a later stage if funds materialize.)

Users thus have every prospect that the telescope will yield a harvest of novel radio sources. Australian radioastronomers recognise that, but so do their colleagues elsewhere. This partly explains why the AT (and its single-dish neighbour at Parkes) has become one of the most cosmopolitan locations in Australia; the flux of visiting users who have won time from peer-review committees is noticeable. (In case people in Canberra should notice, the managers are ready with a list of Australian groups who have won time at national facilities elsewhere.)

The immediate harvest of data apart, one strategic goal of the telescope is to broaden the base of astronomy in Australia, hitherto largely centred at the Universities of Sydney and New South Wales as well as the ANU. Is that yet another sign of the government's seriousness about research? And while it may be calculated that a training in radioastronomy may give a person an interest in fields as different as the theory of plasma physics and the electronic signal processing, there are many places other than Australia in which such a view would be deeply envied.

Parkes, some 50 km away from Narrabri, now has similar status, but an antique air. When construction began in the late 1950s, the instrument was meant to last for two decades or thereabouts (which it has successfully done). Whether intentionally or otherwise, it has the appearance of being something in

between a lighthouse and one of the nineteenth-century maritime buildings around the British coast now converted into museums. There is a solid brick-faced circular tower of uniform diameter, perhaps three domestic stories high. But instead of a lantern on the tope, there is a huge dish antenna.

In the short history of radioastronomy, Parkes has been the chief means of exploring the radio-bright objects in the Southern sky. But the dish has now been successfully reconfigured so that the central parts are accurate enough to be sensitive to millimetre wavelengths. With luck, the observatory should have at least another decade's life ahead of it.

Meanwhile, the telescope has taken its reputation for sobriety in its hands and has signed a contract to allow the private US organization dedicated to the search for extra-terrestrial life (under the banner 'SETI', where the 'I' stands for intelligence) to buy A\$1 million worth of time in the first few months of the year. The strategy is to scan the spectrum accessible at Parkes, looking for inexplicable time-variations of intensity. The ubiquitous Radiophysics Division has apparently built an equipment that will pick out significant events, but nothing telling had been found in roughly a month of operation.

Neither Parkes nor the AT is the heroic peak at these observatories. That distinction goes to the project called SUSI (Sydney University Stellar Interferometer), which is the most tangible embodiment yet of a scheme for measuring the visible diameter of distant stars due to Robert Hanbury Brown, now retired as professor of physics at the University of Sydney, but previously a member of the faculty at Jodrell Bank in Britain.

John Davies, a younger graduate of Jodrell Bank and Hanbury Brown's successor at Sydney, is struggling to make his mentor's dream come true. His equipment is a remarkable construction: a series of 30-cm mirrors at intervals along a 500-metre pipe that connects them all together. The objective is to observe the same star with all the mirrors, but to make the signals from all possible pairs interfere with each other at some central point.

In one sense, it is the problem of using a pair of widely separated mirrors as a range-finder, except that the objective is to measure the diameter, not the distance. Evidently, it is not allowed to convert the signal from each mirror into digital form, for that would throw away phase information. But how to compensate for the different distance between the mirrors in all possible pairs? Davies is working with a kind of toy train system that adds a time delay between one signal and the other. It is an uphill task, especially with a teaching-load in Sydney that cannot be skipped. The Australian Research Council, which has backed the project generously so far, should be willing to open its chequebook again. □

The sensational discovery of X-rays

W. C. Röntgen discovered X-rays about this time a century ago. Within a few months, Becquerel was making the first observations of radioactivity. But the beginnings of the then new physics were not a tidy business.

Würzburg. Later this year, the university in this mediaeval Bavarian city will be celebrating the centenary of the discovery of X-rays by W. C. Röntgen, its most famous son. As with most centenaries the precise date on which the celebrations should take place is a matter for conjecture. Röntgen spent just over a decade in Würzburg before moving to Munich as head of the physical institute there. Before his death, he arranged that his personal papers should be destroyed, so there are no laboratory notebooks through which his successors can search for the first evidence that an electrical discharge in a Crookes tube can sometimes produce what were called actinic rays.

Of course, the landmark really to celebrate would have been the day on which he took the evocative X-ray photograph of the bones of his wife's hand, but it is not known when that may have been. By default and by convention, the day to celebrate is therefore taken to be that at the end of November 1895 when Röntgen read a paper entitled "On a new kind of rays" to the Würzburg scientific society.

Although the Würzburg society published Röntgen's paper promptly, dissemination of news of the discovery began only with the publication of a translation of Röntgen's paper in *Nature* on 23 January 1896. Because the *Nature* archive is in an even more parlous condition than the Röntgen archive, only surmise allows a reconstruction of what happened. Probably Sir Arthur Schuster, then at the University of Manchester, had been given a copy of the paper and urged *Nature* to translate it.

The evidence for that is internal to the published journal. *Nature*'s first reference to Röntgen's work was on 16 January 1896, when, in a short anonymous note, the then assistant editor, Richard Gregory (editor from 1919 to 1942), wrote:

Professor W. C. Röntgen . . . is reported to have discovered that a number of substances which are opaque to visible rays of light, are transparent to certain rays capable of affecting a photographic plate. It is alleged that he has been able to utilise his discovery to photograph metals enclosed in wooden or woollen coverings, and has succeeded in obtaining pictures showing only the bones of living persons The scientific world will look forward with interest to the publication of Professor Röntgen's work.

Gregory was telling readers that something of a scoop would follow.

The English translation of Röntgen's paper appeared on 23 January 1896, with a brief commentary by Schuster. Considering the content of the paper, Röntgen's use of terms was meticulously modest. He used the term "X-rays", but "for the sake of brevity" only. Otherwise, the paper is as sober as anybody could ask: most obstacles are transparent to X-rays, at least when they are sufficiently thin; the transparency of solid objects to X-rays is not simply an inverse function of their mass-density; materials such as aluminium are, mass for mass, more transparent than, say, lead.

Röntgen went on to show that X-rays are neither refracted nor reflected as are rays of ordinary light, but that materials opaque to X-rays also backscatter them, exhibiting what he called "turbidity". He showed that the X-rays were produced by the impact of cathode rays in a discharge tube on the walls of the tube (and that the apparent source could be moved about by means of a magnet).

What could they be, these X-rays? Certainly not ultraviolet light, which was known to be capable of being reflected, refracted and even polarized, in accordance with Maxwell's theory of electromagnetic waves which, among other things, allowed only for the vibrations of the field in which the electric and magnetic fields are mutually orthogonal and also perpendicular to the direction of the ray. So might not X-rays be the 'missing' longitudinal version of Maxwell waves? Röntgen concludes his paper by saying:

I must confess that I have in the course of this research made myself more and more familiar with this thought, and venture to put the opinion forward, while I am quite conscious that the hypothesis advanced still requires a more solid foundation.

In his comment on Röntgen's paper, Schuster acknowledged that this "remarkable discovery will materially affect our views concerning the relation between the ether and matter", but did not like Röntgen's invocation of longitudinal electromagnetic waves. Röntgen, he complained, had not proved that the new rays were not electromagnetic radiation of the kind that Maxwell had foreseen, but of much smaller wavelength than any radiation previously described. Schuster argued that Röntgen's failure to demonstrate interference might be explained by the incoherence of the source, his failure to demonstrate the refraction as a sign that the wavelength of the X-rays is much less than interatomic

distances in real solids. In a perceptive sentence, Schuster added:

If Röntgen's rays contain waves of very small length, the vibrations in the molecules which respond to them would seem to be of a very different order of magnitude from those so far known. Possibly we have here the vibration of the electron within the molecule, instead of that of the molecule carrying with it that of the electron.

Within a few years, *Nature* had published several hundreds of contributions on what it persisted in calling "Röntgen's rays", from Germany, France, Italy and the United States as well as Britain. The central question was the nature of the new phenomenon.

It is possible to appreciate the flavour of the arguments that followed only by recalling that the search for the "luminiferous ether" was still under way, and that J. J. Thomson had not yet proved that "corpuscles of electricity", otherwise electrons, exist with a mass less than the masses of atoms and with negative electric charge. But Thomson's frequent contributions to the debate in the columns of *Nature* vividly illustrate the evolution of his thinking on the nature of electricity.

In practice, there was evidence within about a year of classical diffraction in an account of the transmission of X-rays through a tapering metal slit laid on a photographic plate; the image is more blurred as the width of the slit tapers away to nothing. But the issue was not finally resolved until 1912, with the demonstration of X-ray diffraction from a crystal by Max von Laue.

In this welter of publication, there is a poignant comment. On 20 February 1896, there appears a letter from William Ramsey, who had (with Lord Rayleigh) only recently discovered argon, under the title "Science and morals":

A habit has been growing in recent years. . . . After the announcement of an interesting discovery, a number of persons at once proceed to make further experiments, and to publish their results. To me, it appears fair and courteous, before publication, to request the permission of the original discoverer.

Ramsey was aggrieved that his own work had also been exploited by others, but he also roundly declared that if the practice were to become widespread, "scientific men will provide their laboratories with a good lock". Ramsey's opinions had little influence on *Nature*'s contributors to the X-ray debate. Certainly, modest Röntgen had no complaint. **John Maddox**

Constant compass calibration

James L. Gould

MIGRATING birds have an embarrassing number of compasses, which they calibrate against one another early in life. But the original calibration of the magnetic compass becomes dangerously inaccurate as birds migrate between high and low latitudes. Now, as described on page 230 of this issue¹, Kenneth P. and Mary A. Able have discovered how birds resolve the potentially fatal problem of shifting declinations, and remind us once again how well oiled the machinery of navigation is.

The Ables studied Savannah sparrows, a species that migrates from above the Arctic Circle in Canada to as far south as Central America. Like most migrants, the sparrows fly mainly near dusk or at night; this is probably a good way to avoid overheating and predation. And, like most migrants, they are born with two preprogrammed default directions for migrating. One is magnetic, the other celestial. This redundancy is important because though celestial cues can be very precise, they are unavailable when the sky is overcast.

The young sparrow is presented with two problems. The first is to calibrate its celestial vector against the dusk and night skies. The key factor here is to find true north, which they do over the course of several days by determining the axis of rotation of celestial cues — the pole point in the sky. The rotating cues include the pattern of polarized light in the sky (especially prominent at dusk) and the pattern of stars at night. By itself, however, determining true north from celestial rotation is of limited value. To use either star or polarization patterns under real-world conditions — that is, when clouds may obscure much of the sky — the bird must be able to infer true north from a partial view of the pattern. But because the positions of these patterns change over the course of the evening as the Earth turns, the bird must use its internal clock to calibrate the calibration.

The second problem is more difficult. True north and magnetic north can differ by as much as 90°, especially at high latitudes (see figure). This declination arises because the north magnetic pole is inconveniently located near Bathurst Island in northern Canada, some 1,600 km from the geographic North Pole. For young Savannah sparrows born near Point

Barrow in Alaska, therefore, magnetic north is almost 90° east of true north. Chicks born to parents nesting on the Ungava Peninsula, in northern Quebec at the northeastern mouth of Hudson's Bay, are exposed to an opposite declination — nearly 45° west. So the innate magnetic vector in sparrows must be reoriented by as much as 90° to compensate for the natal declination. Accordingly, young birds calibrate their magnetic

could be done against celestial rotation (as it was initially) or against the magnetic compass. Received wisdom, based on laboratory experiments subjecting migrants to conflicts between magnetic and celestial cues, indicated that nearly all birds are programmed to opt for the second alternative². This suggested to most of us that migrants do not have the leisure to redetermine the axis of celestial rotation *en route*. They must rely on theoretically instantaneous recalibration to magnetic cues.

But as a Savannah sparrow migrates south, it is not just celestial cues that change: the declination shifts inexorably towards 0° as birds leave their nesting areas and move through Canada and the United States. Thus, to rely on magnetic cues at all after leaving the natal grounds seems foolish, and recalibrating celestial cues on the basis of an increasingly mistaken declination value ought to compound the problem.

The Ables wondered if the necessarily artificial conditions and short-term experiments used to test for recalibration might be obscuring some other, more sensible mechanism. They tried a more naturalistic approach, mimicking the rest stops many long-distance migrants make to feed and recover during their arduous travels. By giving the sparrows several days to deal with the changing relationship between celestial and magnetic cues, the Ables found that the birds *will* take the time to redetermine the celestial pole point. In the face of a serious discrepancy between celestial and magnetic cues, they recalibrate the magnetic compass accordingly. Apparently the opposite happens when the birds are given only a day or

two to resolve the discrepancy — under these rushed circumstances, the celestial compasses are reset to the magnetic compass.

It is a relief finally to see the full picture of the innately orchestrated compass-calibration strategy emerge. Now we can understand how sparrows have been managing to do so much better in their migration flights than our models used to allow them to do. □

James L. Gould is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA.



Map showing the range of the Savannah sparrow (red), also indicating why a bird's innate magnetic vector must be reoriented by as much as 90° to compensate for the natal declination.

compass against the celestial pole — true north. At this point, all three compasses are aligned.

Finally, the birds need to recalibrate their celestial compasses and magnetic compass as they grow older and begin to migrate. Why is this later round of recalibration necessary? Consider the celestial cues they are using. As migrants travel south, new star patterns appear over the horizon; the same is true as the seasons change, and new constellations become visible at night. At the same time, the changing day length as the seasons pass alters the relationship between celestial cues (star and polarization patterns), direction and time of day.

To continue using celestial cues, then, the birds must recalibrate these patterns periodically. In theory, this recalibration

1. Able, K. P. & Able, M. A. *Nature* **375**, 230–232 (1995).

2. Able, K. P. in *Orientation in Birds* (ed. Berthold, P.) 166–179 (Birkhäuser, Basel, 1991).

Crumbling canons

Don Cox

It is a little difficult to believe. A measurement of the thermal pressure in nearby interstellar space, averaged over a path of 100 light years or so, is said to shatter our grand conceptions of the interstellar medium. How can the critical connection be made?

The observation and an interpretation of its significance are reported by Bowyer, Lieu, Sidher, Lampton and Knude on page 212 of this issue. Using the Extreme Ultraviolet Explorer satellite, the Infrared Astronomy Satellite and ground-based photometry of stars, they set a lower limit to the thermal pressure in hot gas towards a small nearby cloud of gas, $p = 2nkT \geq 3 \times 10^{-12} \text{ dyn cm}^{-2}$. (This hot gas, at roughly a million kelvin, is thought to occupy a very large region which totally surrounds the Solar System.) The authors note that this pressure is at least a factor of 20 larger than the thermal pressure measured inside another small interstellar cloud, one that at the moment is blowing over the Solar System and is almost certainly immersed in the hotter, higher-pressure environment.

The authors then make their grand leap: the best known and most quoted model for the workings of the interstellar medium, that of McKee and Ostriker in 1977, has as an underlying premise that all components of the medium are in intimate thermal contact and local thermal pressure balance; yet, in the one place that we can measure the thermal pressures of a cloud and its environment, the balance fails miserably. In the authors' view, the model's being untenable at its core is sufficient reason to abandon it in its entirety.

I tried to tell Stu Bowyer, when he first told me of this paper, that the pressure discrepancy between the hot gas and the local wisp of denser gas had been known for a long time, that Ed Jenkins had written about it in 1984, and that Chris McKee himself had been aware of it for a decade. I also tried to tell him that no one believed McKee and Ostriker's model any more; the expectations one develops from the theory simply have too little correspondence to reality. But Bowyer persisted, and brought me round to his point of view. It is that people cannot carry the absence of a reasonable model as their world view. As a result, when pressed for a description about how the interstellar medium works, they will fall back on the McKee and Ostriker orthodoxy. Furthermore, there are those who continue to assert that the model is correct, although perhaps in a somewhat modified form. It is Bowyer's hope that if we are hit on the head with the fact that the underlying

tenets are false, liberation may be possible.

At the time of its publication, I thought McKee and Ostriker's model was one of the most interesting pieces of work I had ever seen. I was very surprised when the subsequent data caused me to reject it as a reasonable synthesis of reality. Several more years passed before I was able to understand how it could be wrong, and I still find some of its concepts useful in looking for a better picture.

There are essentially three tenets of the model. If the supernova rate in the Galaxy is sufficiently high, the explosions should disrupt the interstellar medium into a hot frothy mixture. There is insufficient power to keep all the gas hot, but too much violence to distribute the energy smoothly through the medium. Cold material would find itself in bits and shells or tunnel walls, pervaded by hotter gas of much lower density. This possibility was known before, but McKee and Ostriker proved that the local region of our Galaxy has a sufficient supernova rate to guarantee such disruption. It was not until about two years ago that a way was found around their reasoning, dissolving the certainty but not the possibility. A pervasive phase of hot gas is theoretically possible.

The second tenet was that there is, on average, a local thermal pressure balance between the different components of the medium. This assumption had been around for quite some time, and was loosely based on observations, denser gas usually being colder by about the right amount. But it does not hold in detail, not even between the local wisp and the local hot gas, as Bowyer *et al.* aver. One sometimes assumes that there are important dynamical effects to disrupt the balance temporarily, essentially the point of view suggested by Jenkins, but recent observations of the gas towards the star Capella with the Hubble Space Telescope have found the local wisp to be amazingly free of disturbance. Furthermore, observations of time-varying hydrogen absorption towards pulsars appear to require a significant amount of interstellar gas at extremely high densities. The pressures are probably also very high and with such a large amount of material in this condition, extraordinary dynamics begins to lose its appeal.

One can wonder whether interphase thermal pressure balance is as fundamental to the McKee and Ostriker picture as Bowyer *et al.* hope. My opinion is that it should not be. It was probably assumed as a matter of tradition and simplicity. But it is related to the third tenet, which is fundamental. This is that

mass exchange takes place between the hot gas of expanding supernova disturbances and the colder clouds that it engulfs. Initially clouds are evaporated to raise the density and radiation rate of the hot gas; later the cooling hot gas condenses and is redeposited on the clouds farther out. This mass exchange was invoked in order to have overall mass and energy conservation.

In its purest form, it was imagined that the cloud evaporation took place through thermal conduction to the cloud surfaces. Later, when it became clear that magnetic fields threading the clouds would be swept back over their surfaces to a cometary form by the passing flow of hot gas, proponents began to speak of other ablation mechanisms to avoid the fact that it is difficult for thermal conduction in a plasma to cross a magnetic field.

Why is this related to pressure balance? Without magnetic fields tangential to cloud surfaces, thermal pressure balance should eventually come about. But with a tangential magnetic sheath, it is not required. The measured field strengths in the interstellar medium are in fact of just the right size for such fields to play an important role in equilibrating the interphase pressure. Pressure balance is as useful a concept as ever, but the balance is often dominated by magnetic field rather than thermal pressure. In that case, thermal evaporation of clouds is probably impossible. McKee and Ostriker's pure and original model just cannot work because the field shuts down the first part of the mass exchange. So thermal pressure imbalance, such as that now found between the local hot gas and at least one of the wisps of denser material within it, becomes an important observational indicator of the field strength and surface geometry.

One can inquire whether other ablation mechanisms might save a modified theory; whether the local region might be atypical; how a magnetic field might hold together the extremely high density regions to which I referred earlier (when it is usually used instead to support regions of unusually low thermal pressure); whether real supernova remnants actually show signs of the accumulation of mass from clouds into the hot gas; whether anyone has ever seen a cloud that was being evaporated or ablated by surrounding hot gas; why, if such evaporation occurs, it does not show up in data for the O^{5+} absorption line to the degree expected; or why, looking at other galaxies, we do not see the extensive quantities of hot gas expected in the standard picture to be pervading the disk. These points are all under discussion. So far, in my opinion, there is no evidence that a model like that of McKee and Ostriker can be constructed to resemble the real interstellar medium. The original model certainly does not. ►

So Bowyer was right. With a host of difficulties but no alternative, what do we tell our children when they ask? We lie, I guess. Yes, Virginia, there is a viable model of the interstellar medium, published back in '77.

But let us not end on a depressing note. The interstellar medium is an exciting subject of study, with a great deal known about it. Although we do not know the fundamental reasons why everything is the way it is, we have a partial understanding of many aspects. And to put McKee and Ostriker's model in perspective, it was written in a simpler time when much less was known about how material is distributed, about superbubbles and superdense gas, about the distribution of X-ray-emitting gas, about the magnetic field strength and the almost total insensitivity of its magnitude to gas density (except at very high density), about the vertical structure of the disk and its great scale

height in all but the most quiescent components, and many other aspects. If one wishes to hang on to that theory as somehow relevant, the minimum admission required is not only that it knew nothing of all these things, but that it said nothing about them either. I don't mean that as a hostile criticism; it was a great model. But it is not a synthesis of the current state of knowledge of the interstellar medium and is irrelevant to further useful discussion. To pretend otherwise is to stand in the way of enlightenment. Besides, isn't it more fun for a while to talk about the great mysteries and our feeble attempts to peek behind the tapestry, rather than to describe a whitewash we have thrown over it? □

Don Cox is in the Department of Physics, University of Wisconsin-Madison, 1150 University Avenue, Madison, Wisconsin 53706, USA.

BOTANY

Opening time by degrees

Peter D. Moore

SCIENTIFIC ignorance, especially of matters requiring routine but painstaking screening and recording, is a never-ending source of wonder. Take, for instance, flowering time, about which we know little more than the Victorians. This is one of the main events in the life of most plants, and long-term, environmentally induced changes could well have considerable effects on floral communities and their distribution.

Happily, however, the field naturalist is not extinct, and Richard Fitter has compiled an extensive data set on first flowering dates going back almost four decades. The data are analysed and interpreted by Fitter and others in *Functional Ecology*¹, and the results allow a fairly precise relationship to be detected between the temperatures of preceding months and the first flowering date for a range of plant species. The authors also consider the possible influence of any climate warming on plant communities: their predictable conclusion is that the consequences will be largely unpredictable.

One problem that blights this kind of study is the range of factors that can be involved. There is an underlying genetic basis to flowering requirements, which can vary from one individual to another, as any botanist can observe when particular plants always seem to be the first to flower in a local population. Some plant species also require certain environmental stimuli, such as chilling, before they are able to initiate flower development; and many need a particular, and closely defined, day-length/night-length ratio,

which however can vary even within a particular species.

The cocklebur (*Xanthium strumarium*), for example, is a 'long-day' plant, needing a critically short night-length as a cue for flowering in the spring. But, as a widely-distributed species, it would be inappropriate to exhibit the same requirements wherever it is found, so geographically separated populations vary in their precise demands². In Florida, for example, cockleburs flower once the night-length decreases to about 10 hours, but in northern Michigan the night-length requirement is 8 hours. Selection for the appropriate flowering time has evidently determined the development of a latitudinal gradient of populations with distinct genetic constitutions.

Given the very basic physiological requirements, however, other environmental factors may then determine whether flower development can go ahead, and chief among these is temperature. Its effect upon the date of earliest flowering is apparent from microclimate observations, especially where the microclimates concerned are patchy and diverse, as in alpine situations³. Sheltered, south-facing locations often permit early flowering. But although autecological and physiological information is available for certain (yet surprisingly few) plant species, it would be impossible to predict the effect of weather conditions in a given season on the flowering patterns of whole plant communities. And despite the relative abundance of amateur botanists and taxonomists, there are very few good, long-term systematic

records of flowering dates for specific localities.

Richard Fitter has assembled one such data set by compiling recordings of the first flowering dates (FFDs) of 243 plant species within a few kilometres of his home in Chinnor, Oxfordshire, central England, over the 36-year period between 1954 and 1989. The localized source has the advantage that not only does it eliminate any altitudinal or latitudinal effects; it also lessens the confusion caused by variation between individuals of populations because, at least among perennials, it will be the early-flowering individuals (whether because of genetic or environmental causes) which are likely to be first recorded each year. Apart from the detail and taxonomic reliability of the data, there is the added bonus of a sound set of temperature records for the area over this time period. A team from the University of York has now digested all this information¹, using a range of techniques including stepwise multiple regressions of the mean monthly temperatures of the months preceding flowering in relation to FFD in any given season.

Some species proved much more variable than others in their FFDs. The most variable were annual and ephemeral species, which is what one would expect from plants with opportunistic ecological strategies. When taking advantage of chance disturbance or occasions of reduced competition, a plant that relies heavily on rapid and considerable seed production for its survival cannot afford to be fussy about flowering times. The least variable were those species that flowered later in the year (May to August). These may have been generally less limited by temperature conditions during flower development.

Of the 243 species, all but 24 produced significant results for multiple regressions of FFD on mean monthly temperatures of the months preceding flowering, showing that temperature is a major factor in determining the time of flowering for most species (90 per cent of those recorded). Of the very best fits, three species of *Prunus* fell in the first six plants (*P. cerasifera*, *P. spinosa* and *P. avium*, in that order). Other reliable species included *Sanicula europaea*, *Arum maculatum* and *Oxalis acetosella* (interestingly, all woodland species). The most influential month (not very surprisingly) was February, followed by March, but June and July temperatures were proportionately more important to late-flowering species than would have been expected. An increase in temperature of 1 °C above the norm for any of the four months preceding flowering generally produced an advance in flowering date of between 3 and 4 days. One unexpected outcome of the analysis, however, is the negative influence of the preceding autumn on a large number of species. For

49 plants, a warm autumn led to delayed flowering the following spring; an increase of 1 °C for one of the months falling 6–8 months before flowering retarded flowering by 3–4 days.

The analyses must inevitably lead to questions about the possible impact of climate change, and the simple conclusion is that a warming of climate will certainly affect the flowering dates of most species in such locations (lowland, oceanic, temperate) as those of the survey. The specific effects, however, will depend on the time of year when any such warming is experienced. A uniform rise of, say, 1 °C

would delay the flowering of *P. spinosa* and *Acer pseudoplatanus* by between 10 and 32 days, but would advance the flowering of *Quercus robur* and *Alnus glutinosa* by between 13 and 23 days. It is difficult to predict whether and how such changes would affect the general ecological balance of a community. The plants would presumably remain generally synchronized in their flowering, but insect-pollinated species might find themselves embarrassed by a lack of appropriate vectors just when they most needed them. Extended seasons could also lead to early seed-set and germination, resulting in an

increase in winter mortality.

The Fitter records will by no means answer these questions, but they form a unique contribution of biological grist to the modeller's mill. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

1. Fitter, A. H., Fitter, R. S. R., Harris, I. T. B. & Williamson, M. H. *Funct. Ecol.* **9**, 55–60 (1995).
2. Ray, P. M. & Alexander, W. E. *Am. J. Bot.* **53**, 806–816 (1966).
3. Ellenberg, H. *Vegetation Ecology of Central Europe* 4th edn (Cambridge Univ. Press, 1988).

OBITUARY

Hannes Alfvén (1908–95)

HANNES Alfvén, Nobel Laureate in physics for 1970, died at the age of 86 on 2 April at his home in Djursholm, Sweden. One of the more original figures in physics this century, Alfvén is generally credited with founding the field of magnetohydrodynamics and making it enormously useful to astrophysics.

Because of his radically different approach to physics, Alfvén's ideas were always at variance with more conventional views. He was generally controversial in his presentation of ideas, which were often introduced in an unpolished state that antagonized the workers in the field. In fact, they usually appeared at first sight to be wrong. This was attributed to his lack of comprehension of the standard theory. But more times than not, his ideas prevailed.

Alfvén received his doctorate from the University of Uppsala in 1934, and, after his graduation, became intrigued by the problem of the aurora. To understand better the motion of particles into the Earth from the Sun, he developed the guiding centre approach to the motion of charged particles in a magnetic field. This approach was much simpler than the complicated orbital solutions of Störmer that were formerly used. In fact it is this formulation that underlies all modern work in plasma physics. But because his approach to the aurora was not generally accepted by the geophysical community he was forced to publish his results in a rather obscure journal. It was not until the publication of his book, *Cosmical Electrodynamics*, that the method became widely known.

Using his approach he was the first to discover the adiabatic invariance of the magnetic moment of charged particles in magnetic fields. This is the fundamental concept underlying the mirror machine approach to magnetic fusion,

and the acceleration of cosmic rays. Shortly afterwards, in the early 1940s, Alfvén proposed that the interstellar medium is probably filled with magnetic field lines which confine the cosmic rays. He showed that the currents necessary to sustain this field on cosmic scales were so small that they could easily be produced by the interstellar medium, a fact

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

not appreciated at the time.

I feel that his most significant discovery was that in highly conducting plasmas the magnetic field lines are frozen into the conducting fluid. Thus, they were endowed with a reality that up to this time was discounted by all physicists knowledgeable in electrodynamics. This intuitive idea at one stroke enables one to grasp the complicated behaviour of magnetohydrodynamics. Without it, it would be almost impossible to comprehend the complicated solutions of the three-dimensional fluid equations interacting with electromagnetic fields. For example, with this concept it is easily seen that there must be waves in a magnetohydrodynamic medium that propagate by vibrating the magnetic lines of force as if they were strings. These waves are now known as Alfvén waves.

Perhaps the conversation with Alfvén that I most clearly remember is one in which he described a visit to Chicago just after the Second World War. He told me that he had spent several hours trying to convince Fermi of the reality of these waves, apparently without success. As he was about to leave the United States three weeks later to return to Sweden, he received a postcard from Fermi with the simple message, "You are right!" Apparently this conversation was an important factor in inspiring Fermi's theory of the origin of cosmic rays.

Throughout his life, Alfvén continued to put forth ideas in conflict with conventional scientific thinking. For example, he placed great weight on the importance of double layers, two electric charge layers between which a strong electric field exists. The double layers appeared to me to be absurd, but they are turning out to be important in theories of the aurora and elsewhere.

In spite of his gruff scientific style, Alfvén himself was a warm and kind personality, encouraging and helpful to his colleagues, and particularly to younger scientists. I think it very appropriate that the waves he discovered are named after him. The occurrence of Alfvén waves permeates almost all branches of plasma physics, not merely magnetohydrodynamics. They are mentioned several times a day by nearly every plasma physicist working in the field that Alfvén founded. He is thus immortalized whether he is admired or not. It was my great honour and good fortune to know him, and I will miss arguing with him.

Russell Kulsrud

Russell Kulsrud is in the Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA.

Divine malevolence

Philip B. Allen

EINSTEIN believed that God was subtle but not malicious. He might have altered his view if he had known about high-transition-temperature (high- T_c) cuprate superconductors. Divine malice accounts well for the crystallography and phase stability of these materials, which have plagued efforts to reproduce or interpret experiments, and thus to confirm or refute theories.

Eight years after their discovery, these high- T_c materials are still poorly understood. But at a recent conference in Stanford (*Spectroscopies in Novel Superconductors*, 15–18 March 1995), spectroscopists and theorists mostly agreed on one previously contentious issue: the energy gap in the electronic spectrum, at least for the highest- T_c systems, has 'nodes' at which the excitation energy goes to zero. Another big issue where quarrels are gradually being resolved is that the connection between the super-

conducting cuprates and the insulating antiferromagnetic cuprates, which are crystallographically almost identical, is becoming clearer.

First, some background. Superconductivity is a low-temperature property of many metals. The BCS theory, developed in 1957 by Bardeen, Cooper and Schrieffer, revealed that the secret of the superconducting state is a new form of long-range order created by pairing of electrons. This requires an effective attractive interaction to counteract the Coulomb repulsion. In 'conventional' superconductors this is an indirect interaction, modulated by the ionic lattice (for a rough idea of how this works, consider an electron moving past and attracting an ion: the ion's change in vibration can in turn attract a second electron). Pairing changes the way we have to think about the electron energy levels of the material. Single electrons are Fermi particles which

obey the Pauli exclusion principle: only one of each spin can occupy each level. But when they pair up, the result is a Bose particle which obeys a different law. At low temperatures, Bose particles can all condense into the same state. BCS superconductivity consists of the simultaneous formation and condensation of electron pairs.

The 'pair field' $\psi(r)$ used to describe the superconducting state has multiple roles. It is a sort of two-electron wavefunction attached to each point of the material, and represents not only the form of the electron pair but also the amplitude and phase of the new long-range order. In addition, it is proportional to the gap which opens in the electron spectrum at low temperatures. An ordinary metal has no gap; electronic excited states of arbitrarily low energy are available. In a superconducting material, an energy gap exists only at temperatures below T_c . Above T_c , the pair field disappears, and with it the gap, and the metal becomes a normal ohmic resistor.

Much of this BCS framework also seems to apply to the cuprates. The main issues are the nature of the particles that pair, the pair field they condense into, and the interactions that promote the pairing. It is the form of the pair wavefunction that has particularly been exercising experimenters for the past year. Last summer, V. J. Emery described the continuing debate over the symmetry of the wavefunction, which is highly anisotropic (*Nature* 370, 598; 1994). Since then the evidence already accumulating for d -wave symmetry has strengthened.

Apparently the pair wavefunction $\psi(r)$ varies with direction like the d -wave function $x^2 - y^2$ shown in the figure. This function can also be written as $r^2 \cos(2\theta)$ and changes sign as the angle θ from the x -axis increases by 90° . The Fourier transform of $\psi(r)$ is proportional to the k -space energy gap Δ_k (see box). It has a similar variation, $k_x^2 - k_y^2$. The minimum excitation energy, $|\Delta_k|$, thus has zeros when $k_x = k_y$, at 45° to the k_x axis.

Evidence for a minimum excitation energy $|\Delta|_{\min} \approx 0$ — and thus for nodes — comes from photoemission experiments (J. C. Campuzano, Argonne National Lab.; Z. X. Shen, Stanford Univ.); from Raman scattering (R. Hackl, Meissner Institute; M. V. Klein, Univ. Illinois, Urbana); from quasiparticle tunnelling (Ø. Fischer, Univ. Geneva); and from microwave loss experiments (D. A. Bonn, Univ. British Columbia). The simplest indirect indicator, the low-temperature heat capacity, previously plagued with sample problems, now also agrees (K. A. Moler, Stanford Univ.).

However, although the evidence for nodes in $\text{YBa}_2\text{Cu}_3\text{O}_7$ is conclusive, the interpretation as d -like is not universally accepted. For one thing, the low crystal-

Symmetries and sign change

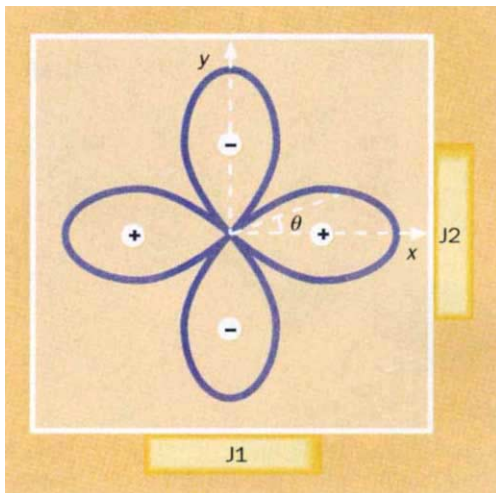
In a normal metal, apart from a periodic modulation, the single electron orbitals are plane waves $\psi(r) = e^{ikr}$ and have energies which vary continuously with wavevector k . The order parameter of a superconductor is the 'pair amplitude'. This is closely related to the wavefunction $\psi(r_1 - r_2)$ of a Cooper pair of electrons at positions r_1 and r_2 . Specifically, the pair wavefunction (apart from a periodic modulation) is

$$\psi(r_1 - r_2) = \sum_k g(k) \exp[ik(r_1 - r_2)]$$

where $g(k)$, the Fourier transform of the pair amplitude, is proportional to the energy gap Δ_k .

The excited states of the superconductor are altered from the normal state. In the normal state it costs energy $|\epsilon_k - \epsilon_F|$ to put an electron into an excited one-electron state, where ϵ_k is the single particle energy spectrum and ϵ_F is the Fermi energy. In the superconducting state, the energy cost is $((\epsilon_k - \epsilon_F)^2 + \Delta_k^2)^{1/2}$. In the normal state the minimum energy cost is zero, whereas in the superconducting state it is instead the smallest value of $|\Delta_k|$.

In an s -wave superconductor, Δ_k is approximately a constant, Δ_0 , independent of k , but in a d -wave state, the gap Δ_k has $d(x^2 - y^2)$ symmetry in k -space: $\Delta_k \propto k_x^2 - k_y^2$. This function has zeros, or 'nodes', when $|k_x| = |k_y|$ (at 45° between the k_x and k_y axes). The corresponding 'real space' pair wavefunction $\psi(r_1 - r_2)$ computed from the equation above will then have d -wave ($x^2 - y^2$) spatial symmetry, with nodes and a sign change upon rotation of 90° , as indicated in the figure.



This sign change is directly measurable by doing an interference experiment using Josephson junctions to transfer Cooper pairs into the d -wave superconductor (from a conventional s -wave superconductor) at one angle, and then back out at an angle rotated by 90° , as illustrated in the picture. The pairs removed at J2 have the sign of their wavefunction changed relative to their sign at the point of insertion J1. In the conventional superconductor which completes the circuit, there is no sign change, and therefore the wavefunction violates the requirement of being 'single-valued' unless the sign change is compensated by a half-integral number of magnetic flux quanta through the loop. This unusual magnetic effect has now been detected (see text) and seems to settle the question of a sign change and nodes in the energy gap.

P. B. A.

lographic symmetry permits mixing of $l=0$ (s -wave) and $l=2$ (d -wave) states. A function like $r^2 \cos(2\phi - \gamma \sin^2 \theta)$ has nodes shifted to greater than $\pm 45^\circ$ for small positive γ . The most powerful $\mathbf{k}$ -resolved spectroscopy, photoemission, measures only $|\Delta_{\mathbf{k}}|$ and cannot either detect a sign change or unambiguously locate a node. Direct evidence for a sign change comes from elegant interference experiments, explained in the box (D. A. Wollman, Univ. Illinois, Urbana; C. C. Tsuei, IBM Yorktown Heights; F. Wellstood, Univ. Maryland; D. A. Brawner, ETH Zurich).

In conventional superconductors, both 'nodes' and antiferromagnetism are harmful to high T_c . Probably the same strong short-range Coulomb repulsion which drives antiferromagnetism is also responsible for the nodes, which may just reflect reluctance of paired electrons to visit simultaneously the same copper site. Thus a subtheme of the Stanford meeting was continued exploration of magnetic cuprate insulators and of the crossover to the metal state as the crystal chemistry

is altered continuously. Spectroscopists (B. O. Wells, Massachusetts Inst. Technol.) and theorists (N. Bulut, Univ. California, Santa Barbara; A. Moreo, Florida State Univ; W. Hanke, Univ. Würzburg; R. B. Laughlin, Stanford Univ.) are now comparing results on the electronic excitations of a model cuprate insulator. NMR (T. Imai, Massachusetts Inst. Technol.) and neutron scattering (G. Aeppli, AT&T Bell Labs; H. Mook, Oak Ridge National Lab.) are discovering that magnetic fluctuations in the metals are much like those in the insulators at high temperature.

The theorists' job is still open. What particles condense and what interaction causes it? How do repulsive interactions which are detrimental to ordinary superconductivity manage to permit (if not build) stronger superconducting correlations with a pair-field containing nodes? Are these nodes compulsory features demanded by symmetry or accidental signatures of a mischievous Deity? Conceivably a modified version of the conventional model could work (A. A. Abrikosov, of

Argonne National Lab., suggested, "God is lazy, and will not create a new theory if He can manage with the old one").

Theories based on ordinary electron quasiparticles interacting through spin fluctuations (D. Pines, Univ. Illinois, Urbana) live happily with d -wave pair-field symmetry. But most theorists suspect that these pictures do not take adequate account of the electron 'correlation' problem which is manifestly present in the insulating state. Theories that omit almost everything except the strong correlations (T. M. Rice, ETH Zurich; D. J. Scalapino, Univ. California, Santa Barbara) suggested d -wave $x^2 - y^2$ pairing well before the experiments, but have still not succeeded in demonstrating a superconducting ground state. Theory is well behind experiment in this field, but has a lot of new data to chew on. □

Philip B. Allen is at the Institut Romand de Recherche Numérique en Physique des Matériaux (IRRMA) IN - Ecublens, 1015 Lausanne, Switzerland.

ECOLOGY

Predicting and producing chaos

Peter Kareiva

ONE of the best reasons for building explicit demographic models is that they can help to predict when subtle changes in the environment might provoke profound changes in population dynamics. On page 227 of this issue¹, Costantino and collaborators provide one of the first examples of a model that actually lives up to this promise.

The authors use a nonlinear demographic model to predict shifts in flour beetle dynamics from stable equilibria to cycles to aperiodic (and even chaotic) fluctuations simply as a result of modest changes in beetle survival rates. They then validate their model by manipulating beetle survival in elegant laboratory experiments and observing exactly the shifts in dynamics that the theory predicts. In particular, beetle fluctuations became highly irregular when adult mortality was experimentally increased (from 73 per cent to 96 per cent). This work is an unusual blend of nonlinear dynamics theory, statistics and experimentation — and the results are of uncommon clarity for ecology.

The theory of nonlinear dynamics, which has chaos theory as its most fashionable expression, is one of this century's more important scientific advances. The surprising richness of behaviour possible from simple nonlinear feedback systems has provided grist for Hollywood and spawned major academic research centres. One would have thought that ecology,

a cradle of chaos theory^{2,3}, would be especially illuminated by studies of nonlinear dynamics; after all, virtually all sensible ecological models are highly nonlinear in a way that admits the possibility of complex dynamics, and observations of population trajectories for animals routinely reveal cycles and wild fluctuations. Yet until recently the study of nonlinear dynamics in ecology has been largely neglected by experimentalists.

The achievements of Costantino *et al.* stem from the unique collaboration involved. Two of the four contributors provided a command of flour beetle biology that led to the design of practical experiments and the formulation of a population model grounded in just the right amount of biological detail (with the key element being density-dependent cannibalism, which provides the nonlinear feedback that drives the dynamics). The third applied his expertise to the analysis of nonlinear dynamical systems to spell out the possibilities for population behaviour. And the fourth member analysed time-series data to estimate model parameters, thereby identifying exactly what sort of experimental alterations of model parameters were needed to produce dramatic shifts in population behaviour. Although a mix of models and experiments in ecology is nothing new, such ventures are rarely accompanied with rigorous parameter estimation, or careful statistics⁴ — it is almost as though theoretical ecologists

have felt that statistics is a subject for empiricists, not modellers.

It is also worth noting that most previous connections between the theory of nonlinear dynamics and natural populations have been aimed at simply establishing that chaos is evident in ecological time series⁵. Only within the past decade have researchers used their understanding of nonlinear dynamics to interpret key features of observed population fluctuations⁶⁻⁹. These reports have all been retrospective, however. In no case were predictions about exact boundaries in parameter space that demarcate different system behaviours tested by manipulating a population's parameters, with the object of seeing whether the dynamics actually did shift between stable equilibria and aperiodic cycles.

To a scientist mired in the mud of the real world, or concerned with only the most pragmatic of issues, it might seem curious to pay so much attention to aperiodic cycles and nonlinear theory. There is a reason, however. Nonlinear dynamics admits the possibility of surprise events occurring in response to subtle modification of a system's parameters. For instance, if we consider the flour

1. Costantino, R. F., Cushing, J. M., Dennis, B. & Desharnais, R. A. *Nature* **375**, 227–230 (1995).
2. May, R. M. *Science* **196**, 643–647 (1974).
3. Gleick, J. *Chaos: Making a New Science* (Penguin, New York, 1987).
4. Ludwig, D. *Ecology* **76**, 357 (1995).
5. Hastings, A. *et al.* *A. Rev. Ecol. Syst.* **24**, 1–33 (1993).
6. Murdoch, W. W. & McCauley, E. *Nature* **316**, 628–630 (1985).
7. Grenfell, B. J., Price, O. J., Albon, S. D. & Clutton-Brock, T. H. *Nature* **355**, 823–826 (1992).
8. Hanski, I., Turchin, P., Korpimäki, E. & Henttonen, H. *Nature* **364**, 232–235 (1993).
9. Tilman, D. & Wedin, D. *Nature* **353**, 653–655 (1991).

beetle populations studied by Costantino *et al.*, we might think that an increase in mortality of adult beetles caused by a pesticide application would simply reduce the populations affected. But an apt non-linear model indicates that the increased mortality could actually cause beetle populations to shift from a stable equilibrium to aperiodic fluctuations. In different nonlinear models for other organisms it could turn out that subtle decreases in mortality or increases in reproduction could cause chaotic dynamics.

The point is that simply documenting

direct demographic effects of environmental change or environmental stress has little predictive power. Only with a solid understanding of underlying dynamics can ecologists ever hope to anticipate the consequences of environmental perturbation. Costantino *et al.* show us how the marriage of nonlinear models and experiments can help to accomplish the task. □

Peter Kareiva is in the Department of Zoology, University of Washington, Seattle, Washington 98195, USA.

repeats of this motif or variants thereof), and the addition of flanking sequences, have generated response elements which are highly specific for a given receptor species^{2,3} (*a* in the figure). Nonetheless, both receptors and response elements enjoy considerable promiscuity — to an extent that NGFI-B, previously praised as the prototypic 'single' among nuclear receptors², is now reported to heterodimerize with RXR on DR5 (ref. 4). Amazingly, this promiscuity follows precise rules and this is where the story has become exciting.

Two main points had to be resolved: the mechanisms generating the response-element repertoires characteristic for a homo- or heterodimer; and, in view of the modular structure of nuclear receptors, the structural elements involved. It was already textbook knowledge that the P-box (*b* in the figure) of steroid hormone receptor DNA-binding domains is part of the DNA-recognition helix able to distinguish between an oestrogen and a glucocorticoid response element due to its specific base contacts. It was also known that the D-box, together with additional residues of the CII zinc-finger motif, forms a symmetrical dimerization interface in the DNA-binding domains of steroid and thyroid hormone receptors which sets the different spacings of their cognate palindromic response elements (refs 2, 5 and references therein). But it was unclear how the various heterodimeric RXR complexes could recognize their distinct direct-repeat response-element repertoires (blue box, *a* in the figure). Unlike

NUCLEAR RECEPTORS

How to finger DNA

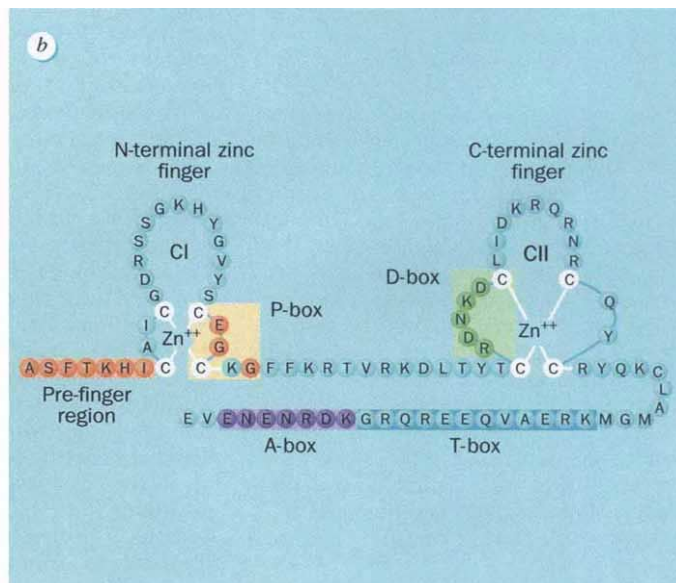
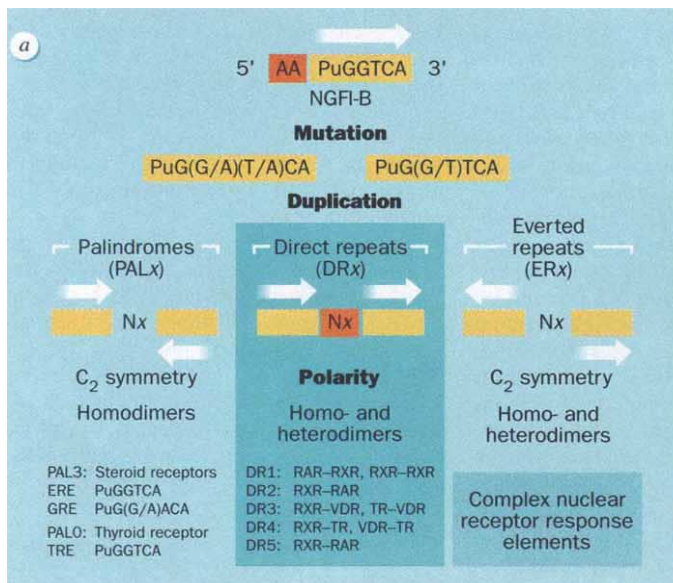
Hinrich Gronemeyer and Dino Moras

A CONVERGENCE of biochemical and structural analyses, aimed at understanding the molecular mechanisms involved in DNA recognition by nuclear receptors, has now culminated in the publication of the first three-dimensional structure of a receptor heterodimer bound to its cognate response element.

On page 203 of this issue¹, Rastinejad *et al.* describe the crystal structure of the heterodimer formed between the DNA-binding domains of the retinoid X (RXR) and thyroid hormone (TR) receptors and their cognate response element, a direct repeat of the canonical 5'-PuGGTCA recognition motif spaced by four nucleotides (DR4). This is the first nuclear

receptor heterodimer and the second heterodimeric DNA complex, after c-Jun/c-Fos, that has been crystallized. The report also provides a lesson about the elegant principles of molecular recognition which create surfaces precisely designed to accommodate homo- or heterodimeric complexes, or both, on symmetrical or non-symmetrical response elements. DNA recognition is the first step which specifies the various ligand-dependent gene programmes, so this lesson is worth learning.

All nuclear receptors recognize derivatives of a canonical hexameric DNA core motif. But mutation and duplication (leading to distinct relative orientations of



Response elements of nuclear receptors and regions within their DNA-binding domains involved in response-element selection. *a*, The canonical core recognition sequence is PuGGTCA which, together with two 5' As, is a response element of the orphan receptor NGFI-B. Duplication of the core sequence generates palindromes (PALx) and everted repeats (ERx), and polar direct repeats (DRx), with *x* base pairs separating the two half-sites. PAL3 is an oestrogen response element (ERE); PAL0 is a thyroid hormone response element (TRE). A mutation in the core sequence results in PAL3 response elements recognized by the

glucocorticoid receptor (GRE) and other hormone receptors. Whereas PALs bind homodimers, DRs can bind homo- or heterodimers with the specificities indicated. RAR/RXR, retinoid receptors: VDR, vitamin D receptor; TR, thyroid hormone receptor. *b*, The RXR DNA-binding domain, highlighting the regions involved in response-element selection. The P-box is part of the DNA-recognition helix¹, and swapping of red-circled residues switches ERE/GRE recognition. The D-box is responsible for PAL3/PAL0 selection by oestrogen or thyroid hormone receptors. (For details see refs 1–3, 6 and references therein.)

previously crystallized symmetrical complexes, direct repeats exhibit polarity and should thus generate anisotropic receptor-DNA complexes. In the environment of a directional promoter, their properties may be distinct from those of symmetrical complexes.

Biochemical studies paved the way. They showed not only that the DNA-binding domains of RAR, RXR and TR dictated the distinct response-element repertoires of the various homo- and heterodimers, but also that the polarity of heterodimers on direct-repeat elements depended on both the DNA-binding domain and the direct repeat concerned — that is, 5'-RXR-RAR on DR2 and DR5; 5'-RXR-TR on DR4; 5'-RAR-RXR on DR1 (refs 6, 7 and references therein). From these results, two mechanisms were proposed to generate specificity and polarity: dimerization of the DNA-binding domain, resulting in the observed cooperative DNA binding to cognate elements; and steric hindrance, which would prevent the concomitant binding of two monomers. By swapping techniques, it was possible to define particular regions within these receptor DNA-binding domains which forced a chimaeric domain to adopt a distinct DNA-binding specificity⁶.

The beauty of crystallography is that it leaves no doubt. The three-dimensional structure of the RXR-TR-DR4 complex solved by Rastinejad *et al.*¹ impressively confirms what the biochemical data only hinted at.

First, the polarity: the RXR DNA-binding domain does indeed occupy the 5' position. Second, the asymmetric dimerization interfaces: the crystallographic data show how a dimerization interface is formed to fit precisely to a given direct-repeat spacing. The RXR surface is generated by CII residues, including a D-box arginine, and the TR side contributes a surface comprising residues of the prefinger, CI and T-box (*b* in the figure).

Third, two principles generate direct-repeat spacer selectivity. One is steric hindrance: a helix which encompasses the TR A-box and forms interesting backbone and minor groove interactions is the key element which precludes the formation of RXR-TR on direct repeats spaced by less than four base pairs. Because of this helix, TR accounts for two-thirds of the DNA contacts within the complex. The other principle is that dimer interfaces can only form on the cognate direct-repeat elements. Rastinejad and colleagues' modelling of RXR-RAR on DR5 is in good agreement with studies determining the sequences of DNA-binding domains that are specifically required for the cooperative binding to a given direct-repeat element⁶. It is now clear that the second strong dimerization function found in the ligand-binding domain has no selective

power for response-element recognition, but only stabilizes the complex. This is most obvious from swapping experiments, which demonstrate that the response-element recognition by the full-length receptors is dependent only on the elements specific to the corresponding dimer interfaces on their DNA-binding domains (C. Zechel and H. Gronemeyer, unpublished data).

Other points of interest emerge from the new study¹. The carboxy-terminal helix, observed by nuclear magnetic resonance in the RXR DNA-binding domain⁸, is not visible due to crystallographic disorder, underlining a structural difference between 5' and 3' binding domains; this helix may play a role in other homo- or heterodimeric complexes of RXR. And water-mediated hydrogen bonds account for roughly half of the interactions between DNA and its binding domain. In this respect the involvement of water molecules in facilitating the interactions with non-cognate response elements is highly informative⁹.

What next? Certainly, the modelling studies¹ for RXR and the vitamin D receptor on DR3 and RXR-RAR on DR5 should be confirmed. What could be even more interesting will be to image the two other retinoid receptor complexes, RXR-RAR on DR2 and RAR-RXR on DR1 (refs 6, 7). In particular, the RAR-RXR interface on DR1, which possibly com-

prises the RXR T-box and RAR CII, and gives rise to the opposite polarity, warrants especial attention. Even more important, there is evidence that the actual sequence of the core motif and the receptor polarity may affect ligand binding and/or the transactivation potential of the AF-2 functions of the heterodimeric partners^{7,10}. Deciphering the structural basis and the consequences of such allosteric effects between DNA- and ligand-binding domains, as well as better understanding of non-cognate DNA interactions on most natural response elements^{5,9}, will give us a glimpse of the complexity of DNA recognition by transcriptional enhancer factors. □

Hinrich Gronemeyer and Dino Moras are in the Institut de Génétique et de Biologie Moléculaire et Cellulaire, Université Louis Pasteur, 1 rue Laurent Fries, Illkirch, C.U. de Strasbourg, France.

1. Rastinejad, F., Perlmann, T., Evans, R. M. & Sigler, P. B. *Nature* **375**, 203–211 (1995).
2. Glass, C. *Endocrine Rev.* **15**, 391–407 (1994).
3. Giguère, V. *Endocrine Rev.* **15**, 61–79 (1994).
4. Perlmann, T. & Jansson, L. *Genes Dev.* **9**, 769–782 (1995).
5. Schwabe, J. W. R., Chapman, L. & Rhodes, D. *Structure* **3**, 201–213 (1995).
6. Zechel, C. *et al.* *EMBO J.* **13**, 1425–1433 (1994).
7. Kurokawa, R. *et al.* *Nature* **371**, 528–531 (1994).
8. Lee, M. S., Kliever, S. A., Provencal, J., Wright, P. E. & Evans, R. M. *Science* **260**, 1117–1121 (1993).
9. Gewirth, D. T. & Sigler, P. B. *Nature struct. Biol.* (in the press).
10. Durand, B. *et al.* *EMBO J.* **13**, 5370–5382 (1994).

EARTHQUAKES

One shock leads to another

Tom Henyey

ON page 221 of this issue, Harris, Simpson and Reasenber¹ examine possible linkages between nearby earthquakes in southern California. The authors find that in the first 18 months after an earthquake of magnitude 5 or more, any subsequent large earthquake (also of magnitude 5 or more) that happened nearby almost always ruptured a fault that had been loaded towards failure through changes caused to the static stress field by the first event. If this is the case, what might it mean for earthquake hazard assessment?

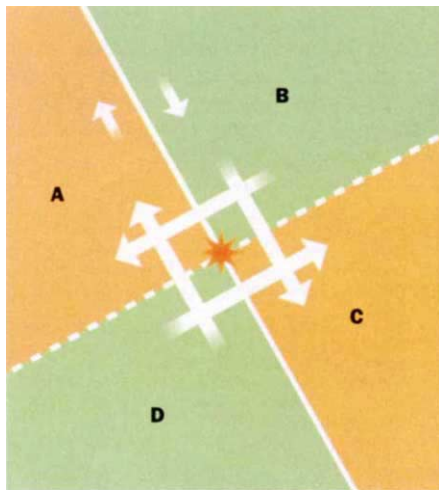
An earthquake can be modelled as a finite dislocation in an elastic medium and represented by a double-couple change in stress (see figure on following page). Although the accumulated shear stresses along the dislocation surface (fault) are released by the earthquake, the crust adjacent to the fault in quadrants A and C in the figure is compressed, whereas that in quadrants B and D is dilated. According to the theory of elasticity, this redistribution of the static stress field will change both the normal and shear stresses acting across any plane of arbitrary

orientation in any of these quadrants. If one such plane happens to be a nearby active fault, and its normal stress is decreased and/or its shear stress increased in its preferred direction of slip, the fault will be moved closer to failure. In a population of faults with random orientations, there will always be some that are moved closer to failure, and as one or more of these faults may be on the brink of failure already, the incremental change in stress may be sufficient to induce an earthquake in the near future.

Following the magnitude 7.3 earthquake that struck Landers, California, in 1992, several groups^{2–4} studied, retrospectively, the effects of stress redistribution from earlier earthquakes on the impending Landers rupture, as well as possible stress loading from Landers on other nearby faults. The Landers fault system was found to have been moved closer to failure by earlier earthquakes in the nearby Imperial and Coachella Valleys to the south. Moreover, the causal fault of the 1992 Big Bear earthquake, the largest aftershock in the Landers sequence (its

magnitude was 6.2), was also discovered to have been moved closer to failure by the Landers mainshock.

These discoveries of possible 'earthquake triggering' prompted Harris *et al.*¹ to study 44 earthquakes of magnitude 5 or more that occurred in southern California between 1968 and 1993. They examined the stress interactions (defined by the authors as a change in 'Coulomb failure stress' or ΔCFS) between the 946 unique pairs of events. Of the 946 pairs, 907 were eliminated because the stress interactions were less than 0.1 bar, and another 12 pairs were eliminated because the earthquakes in each pair were within 5 km of



Stretch and squeeze — double-couple representation of an earthquake on a finite right-lateral strike-slip fault. The solid line represents a fault trace, and the star indicates the earthquake epicentre. The regions A and C are compressed, whereas regions B and D are dilated.

one another, and calculations of the stress interactions were considered unreliable. Of the remaining 27 pairs (all with hypocentral separations less than 40 km), 15 of 16 had positive ΔCFS — that is, the first earthquake of the pair moved the fault of the second earthquake closer to failure — and occurred with time delays of one and a half years or less. After the first 18 months there were roughly equal numbers of positive and negative ΔCFS pairs.

After rejecting any systematic bias in the data, the authors conclude that faults that are within 40 km of an earthquake and oriented to produce a positive ΔCFS are more likely to generate an earthquake within the next year and a half than faults oriented to produce a negative ΔCFS . They further suggest that these data are evidence for earthquake-triggering by static stress changes at magnitude 5 or more and that simple models of Coulomb failure stress can mark regions (or, by inference, faults) of increased seismic hazard following a southern California earthquake.

Of course, assuming that earthquakes are linked through static stress changes as Harris *et al.* suggest, the key questions are: do we have a potential earthquake predictor, and if so, how effective is it? The answer is mixed. If the orientations of active fault planes within a seismotectonic region such as southern California can be determined from surface geological mapping or, in the case of buried faults, inferred on the basis of subsurface geological data, then the change in Coulomb failure stress due to an earthquake within the region can be calculated for each such fault immediately after the event. Faults or clusters of faults with similar orientations having a positive ΔCFS of more than 0.1 bar can be singled out as candidates for subsequent earthquakes or aftershocks.

In practice, it is not that simple, for a couple of reasons. First, because of the large number of active faults with similar and conjugate orientations in southern California, the method is not particularly restrictive. Perhaps a half to two-thirds of the area around an earthquake in southern California might be eliminated from the zone likely to experience a second earthquake within 18 months, but that still leaves hundreds of square kilometres and many faults as possible candidates. And second, many large aftershocks, and even some mainshocks, occur on unrecognized or minor faults for which it is either impossible or impractical to calculate the ΔCFS following a nearby earthquake. Although some of these faults may be automatically included as candidates for a subsequent earthquake because they are associated and aligned with larger, more recognizable faults with positive ΔCFS , earthquake locations cannot yet be pinpointed on them by mapping static stress change.

So although the results of Harris and co-workers are scientifically intriguing and may shed some light on why earthquakes often tend to cluster in time and space⁵, the static stress change method will need refinement before becoming economically and socially viable for hazard mitigation, and before being accepted by emergency officials and the public at large. This conclusion should be viewed as a challenge to seismologists and not as a discouragement. □

Tom Henyey is in the Southern California Earthquake Center, Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA.

1. Harris, R. A., Simpson, R. W. & Reasenberg, P. A. *Nature* **375**, 221–224 (1995).
2. Harris, R. A. & Simpson, R. W. *Nature* **360**, 251–254 (1992).
3. Jaume, S. C. & Sykes, L. R. *Science* **258**, 1325–1328 (1992).
4. Stein, R. S., King, G. C. P. & Lin, J. *Science* **258**, 1328–1332 (1992).
5. Kagan, Y. Y. & Jackson, D. D. *J. geophys. Res.* **96**, 21419–21431 (1991).

Dry as dust

Most liquids wet most solids. Their short-range molecular attractions are comparable and fairly feeble, so spreading is easy. Daedalus now plans to change this. He remarks that water often fails to wet a dusty surface. It forms droplets which, covered with dust themselves, roll around without wetting. So he is inventing a dustier anti-wetting dust. Cunningly, it is magnetic.

Magnetic liquids, essentially suspensions of fine magnetic particles, already find many uses. Chemically absorbent varieties can even scavenge specific molecules from a solution, so that they can be segregated by a magnet. DREADCO's chemists are now extending this technology. They are devising fine magnetized particles which are partially hydrophobic, like detergent molecules. As a result, they are surface active. When dispersed in water, they collect preferentially on its surface with their hydrophobic regions facing outwards, but attracting each other laterally. In effect, they mimic the surface-tension mechanism of normal liquids, but at up to a thousand times the intensity. They will cover any liquid with a tough, tenacious, non-wetting skin.

Like a good detergent, DREADCO's 'Dry Water' will be very economical. A few parts per thousand will make water and most other liquids utterly non-wetting. The obvious market is as a final rinsing additive for domestic washing. By drying the water, it will allow clothes, crockery and even bathers to emerge bone-dry from the wash. But Daedalus fears that the inky blackness of the fine magnetic-particle suspension will discourage the domestic purchaser. Instead, he plans to add the product to industrial canals and waterways, many of which are fairly black already.

Boats and barges will cease to be wetted by the water. They will experience much lower drag, and will suffer less from small leaks (which dried water will not easily penetrate). The vessels will even be pushed slightly upwards out of the water. They will ride higher, usefully reducing their draught.

On a smaller scale, blobs of dried water will act as flexible, almost frictionless ball-bearings for many mechanisms. Dry Water will also allow microchemists to pour and handle small liquid samples without loss, like so many blobs of mercury. And the fun of poking little beads of fluid about, which an older generation used to enjoy with mercury, will return to our teaching laboratories. Modern chemophobic students, who flee in terror if a thermometer breaks, will have the same innocent fun with dried water.

David Jones

HIV results in the frame

The following is a selection of the correspondence received about two papers published on 12 January on the dynamics of HIV infection *in vivo*.

Results confirmed

SIR — Elegant studies using early, frequently sampled patients on nevirapine and protease inhibitors (ABT-538 and L-735,524) have now made it possible to calculate the rapid turnover of HIV-1 virions in infected patients^{1,2}: the rate of plasma retroviral decay revealed a half-time of about 2 days for free virus particles or virus-producing cells, indicating that about 30% of the plasma virus population in infected patients is replaced every day. These observations have important implications for our understanding of HIV pathogenesis, and here, using a similar analysis in our previously analysed cohort of 11 AZT-treated patients³⁻⁵, we confirm the earlier findings^{1,2}.

Our patients were homosexual men with CDC group IV disease, who have

Hence the estimated half-time, being a function of the survival of virus particles in the serum and the decay of virus-producing cells, can represent only the upper limit of viral half-time, assuming immediate and absolute inhibition of viral synthesis, which would seem implausible.

The table shows the analysis of our patients. For each patient the initial viral load (in copies per ml), the minimum virus load (and the day when it was reached), the rate of decline, the daily turnover, the half-life time and the data points used for the fit of the exponential decay curve are shown. The average rate of virus decline is 0.46 ± 0.31 (days), the daily turnover is $35 \pm 17\%$, and the half-life time is 1.9 ± 1.1 days. (For these averages we leave out patient 1, who never achieved a 50% decline in serum viral load and whose viral half-life time is about 9 standard deviations

DYNAMICS OF HIV INFECTION

Patient	Initial virus load	Minimum virus load	Rate of decline	Daily turnover	Half-life time	Data points used for fit
1	2,300	1,480 (7)	0.06	0.06	11.0	2
2	30,000	3,000 (8)	0.27	0.24	2.6	3
3	18,000	80 (7)	1.10	0.67	0.6	3
4	320	40 (7)	0.30	0.26	2.3	2
5	420	80 (6)	0.35	0.30	2.0	3
6	1,680	140 (14)	0.34	0.29	2.0	3
7	2,380	640 (6)	0.17	0.16	4.0	3
8	8,480	140 (8)	0.35	0.30	1.0	4
9	11,760	140 (8)	0.63	0.46	1.1	4
10	25,600	440 (8)	0.88	0.58	0.8	3
11	1,280	220 (7)	0.25	0.22	2.8	2
Mean $\pm$ 1 s.d. (with patient 1)			0.43 ± 0.31	0.32 ± 0.18	2.7 ± 2.9	
Mean $\pm$ 1 s.d. (without patient 1)			0.46 ± 0.31	0.34 ± 0.17	1.9 ± 1.1	

had no prior AZT therapy⁵. Serum HIV-1 RNA only from enveloped virions was measured with an assay previously described⁶. After starting treatment, serum samples were taken every 2–3 days.

Virus levels declined rapidly and reached a nadir (or inflexion point) after 6–14 days of therapy. AZT prevents the infection of new cells, but not the production of virus from cells already infected. Thus we expect serum virus to decline according to $v(t) = v(0)[ue^{-at} - ae^{-ut}]/u - a$ ¹. Here $v(t)$ denotes plasma virus at day t after treatment has started; $v(0)$ is the initial viral load; a and u describe the decay of the virus-producing cells and free serum virus, respectively. We do not have enough data points to fit this function, but we approximate the compound decay by a simple exponential decline. It is important to note that data on serum HIV-1 RNA do not allow us to determine which of the two decay processes is rate-determining,

higher than the mean value.) This is in excellent agreement with earlier observations and the results obtained by Wei *et al.*¹ and Ho *et al.*². These data confirm the high level of virus production and rapid turnover of HIV particles in infected patients.

Martin A. Nowak, Sebastian Bonhoeffer

Department of Zoology,
University of Oxford,
Oxford OX1 3PS, UK

Clive Loveday, Peter Balfe,

Malcolm Semple, Steve Kaye,

Melinda Tenant-Flowers, Richard Tedder

Department of Virology,
University College London Medical School,
London W1P 6DB, UK

1. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).

2. Ho, D. D. *et al.* *Nature* **373**, 123–126 (1995).

3. Loveday, C. *et al.* 3rd International HIV Drug Resistance Workshop Abstr. 61, Hawaii (1994).

4. Loveday, C. in *The Association of Clinical Pathologists Yearbook 1995* (ed. Haeney, M. R.) Abstr. 6 (Ass. Clin. Path., London, 1994).

5. Loveday C. *et al.* *Lancet* **345**, 820–824 (1995).

6. Semple, M. G. *et al.* *J. virol. Meth.* **41**, 167–180 (1993).

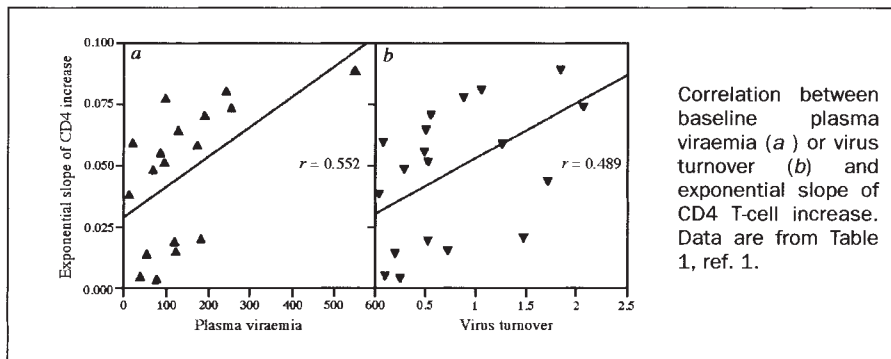
CD4⁺ cell turnover

SIR — The provocative papers by Ho *et al.*¹ and Wei *et al.*² both conclude that antiviral therapy of HIV infection results in a rapid turnover of CD4 T cells, leading to the net production of about 2×10^9 CD4 T cells each day. Three explanations could account for the rapid increase in CD4 T cells in the first weeks after initiation of therapy: (1) new production from thymic progenitors; (2) proliferation of peripheral CD4 T cells; or (3) mobilization of CD4 T cells from infected lymphoid tissue. New production of CD4 T cells in the thymus of adults with chronic HIV infection seems highly unlikely³, so the competing hypotheses are cell proliferation versus cell mobilization.

Both papers conclude that there is a high rate of CD4 T-cell proliferation which, in the absence of antiviral treatment, is balanced by an equally high rate of HIV-induced CD4 T-cell death — the mobilization of CD4 T cells is not considered. The data shown in Fig. 2b of ref. 1 are informative for deciding between proliferation versus mobilization of CD4 T cells. The rate of CD4 T-cell expansion might be expected to be independent of or positively correlated with pre-therapy CD4 T-cell counts, because patients with the lowest starting CD4 T-cell count should have fewer proliferating T cells as well as destruction of lymph-node architecture^{4,5}. However, the patients with the lowest starting CD4 T-cell count actually show the highest rate of CD4 T-cell recovery (Fig. 2b of ref. 1).

An alternative possibility is that a higher viral load is correlated with more trapping of CD4 T cells in lymphoid tissues, and that effective antiviral therapy liberates these cells into the peripheral circulation. Were this explanation to be true, then replottting Fig. 2b of ref. 1 as the exponential slope of CD4 increase versus the baseline plasma viraemia should show as significant correlation as the original figure. The recalculated figure is shown here (a). The correlation coefficient is +0.55, which is statistically identical to the original figure plotting the slope of CD4 T-cell change versus starting CD4 number ($r = -0.57$). If the rate of virus turnover is plotted against the slope of CD4 T-cell increase (b in the figure), a positive correlation is also seen, so both a high starting virus load and a large drop in virus load following therapy predict more rapid CD4 T-cell recovery, and their effects cannot be distinguished from a low initial CD4 T-cell count. It is thus equally likely that the starting viral load and the magnitude of the antiviral effect determine the ability of CD4 T cells to reappear in the peripheral circulation, which is consistent with the mobilization hypothesis.

In addition, the 22-fold variation in the



exponential slope of "CD4 lymphocyte turnover" (0.004–0.088) shown in Table 1 of ref. 1 is inconsistent with replacement of CD4 T cells solely by rapid cell division, as the exponential slope should be relatively constant (reflecting mainly the relatively minor variations in cell-cycle time). Examination of lymph nodes from both HIV-infected progressors⁴ and non-progressors⁵ provides evidence for B-cell proliferation in germinal centres, but little evidence for a pool of proliferating CD4 T cells. In conclusion, a large part of the increase in CD4 T cells seen after acute antiviral therapy is probably due to reduced trapping in lymphoid tissue and a short-term reappearance of cells in the peripheral circulation.

Donald E. Mosier

Department of Immunology-IMM7,
The Scripps Research Institute,
La Jolla, California 92037, USA

1. Ho, D. D. *et al. Nature* **373**, 123–126 (1995).
2. Wei, X. *et al. Nature* **373**, 117–122 (1995).
3. Mackall, C. *et al. Nature* **332**, 143–149 (1995).
4. Embretson, J. *et al. Nature* **362**, 359–362 (1993).
5. Pantaleo, G. *et al. New Engl. J. Med.* **332**, 209–216 (1995).

SIR — Wei *et al.*¹ and Ho *et al.*² report that administering inhibitors of viral replication to HIV-infected patients causes a considerable, transient increase in CD4⁺ cell counts in the blood. From this, the authors conclude that the production of CD4⁺ cells during HIV infection is very rapid and sufficient to replace about 5% of the CD4⁺ cell pool each day. Wei *et al.* then speculate that it may be possible to achieve "... successful immunological reconstitution even in late-stage disease if effective control of viral replication can be sustained." Not surprisingly, this idea has elicited wide interest.

But the data on CD4⁺ cell dynamics^{1,2} are based solely on counting numbers of CD4⁺ cells in blood: in neither study did the authors provide direct evidence on the rate of production (division) of CD4⁺ cells. The tacit assumption in both studies is that finding an increase in CD4⁺ cells in the blood is indicative of an increased production of these cells. In making this assumption, however, the authors ignore the possibility that their data simply reflect an alteration in lymphocyte migration.

Like CD8⁺ cells and B cells, most

CD4⁺ cells are long-lived cells which recirculate continuously between blood and lymph via the lymphoid tissues^{3–6}. But it is well known that lymphocyte migration is labile and can be radically altered by exposure to various infectious agents or to stress⁴; rapid alterations in lymphocyte migration can apply to resting cells and be associated with little or no change in the rate of lymphocyte turnover. Because viral infections often impede lymphocyte migration through the lymphoid tissues (the phenomenon of "trapping")^{3,4}, the transient increase in CD4⁺ cell counts in the blood of HIV patients given antiviral agents could be a reflection of enhanced lymphocyte recirculation (decreased trapping) due to the reduction in the viral load. If so, most of the CD4⁺ cells would be expected to display a resting rather than an activated phenotype. This is easily tested.

To reiterate, we take issue with the view that the data of refs 1 and 2 demonstrate that CD4⁺ cells have a high rate of self renewal. In our opinion the transient elevation of CD4⁺ cells seen in the blood could be an epiphenomenon: the reduction in the viral load induced by the antiviral agents reduces trapping of CD4⁺ cells and thereby facilitates recirculation of these cells.

Jonathan Sprent

David Tough

Department of Immunology,
Scripps Research Institute,
La Jolla, California 92037, USA

1. Wei, X. *et al. Nature* **373**, 117–122 (1994).
2. Ho, D.D. *et al. Nature* **373**, 123–126 (1994).
3. Ford, W.L. *Prog. Allergy* **19**, 1–59 (1975).
4. Sprent, J. *The Lymphocyte: Structure and Function* (ed. Marchalonis, J. J.) 43–112 (Dekker, New York, 1977).
5. Sprent, J. *Curr. Opin. Immun.* **5**, 433–438 (1993).
6. Sprent, J. & Tough, D. *J. exp. Med.* **179**, 1127–1135 (1994).

SIR — Ho *et al.*¹ and Wei *et al.*² report that levels of cell-free virus in the plasma of HIV-1-infected patients fell approximately 100-fold within 2–4 weeks of the start of antiviral therapy, and was accompanied by a rise (approximately 200 cells per μ l) of CD4⁺ lymphocytes. The authors assumed that the rise in CD4⁺ cell number observed equalled the number of cells that were not infected and killed by HIV-1 as a consequence of the effective therapy, and calculated that HIV-1 infected and killed approximately

4×10^7 CD4⁺ peripheral blood lymphocytes each day. They also proposed that the measured rise in circulating CD4⁺ cells reflected an increase in the total number of CD4⁺ lymphocytes in all body compartments of an infected individual. Because only 2% of lymphocytes are present in the peripheral blood, the authors multiplied the CD4⁺ cell increase by 50 and calculated that an HIV-1 AIDS patient can replace approximately 2×10^9 cells each day. We believe that these analyses grossly overestimate the number of the CD4⁺ cells killed by the virus and the regenerative capacity of the immune system in AIDS patients.

We have monitored wild-type and mutant HIV-1 infection kinetics in a variety of tissue culture systems and have derived infection rate constants that allow two critical parameters of the virus life cycle to be calculated: (1) the number of infectious virus particles (n) produced during a single cycle of replication; and (2) the time (t_i) required to complete one cycle of infection³. During a single cycle, each infected cell generated approximately 10^5 physical particles and 10–100 infectious particles, depending on the particular isolate used.

The t_i calculated for several independent tissue-culture infections (3–4 days) is strikingly similar to the reported half-life (1.5–2 days) of cell-free virus particles (measured by RNA PCR) or HIV-1-producing cells in individuals responding to potent antiviral agents^{1,2}. We have also determined the *in vivo* infection rate constants and n values from published reports of individuals undergoing a primary HIV-1 infection or AIDS patients receiving potent antiviral drugs^{2,4–6}. The rate constant for the exponential increase of drug-resistant viral RNA in the plasma of AIDS patients was reported to be 0.27 (ref. 2). We calculated a very similar rate constant (approximately 0.3) for patients undergoing a primary HIV-1 infection^{4–6}. Assuming that the length of a single cycle *in vivo* is also 3 days, the value of n under these conditions would be 2.5 ($n = \exp(kt_i) = \exp(0.3 \times 3)$). Thus 2–3 infectious HIV-1 particles, on average, are released from each cell in the two types of *in vivo* infections.

An independent quantitative study, using an end-point dilution approach, reported that one infectious particle corresponded to about 6×10^4 physical particles in HIV-1-infected individuals⁷. Thus, in both tissue culture and *in vivo* infections, an infected cell produces approximately 10^5 physical particles and far fewer infectious virions.

The two recent *Nature* reports^{1,2} estimated the number of newly infected/replaced CD4⁺ cells in the peripheral blood of AIDS patients to be 4×10^7 per day. As noted above, this number of infected cells should gen-

erate 10^{12} – 10^{13} physical virus particles (4×10^7 cells $\times 10^5$ virions released per cell). However, the plasma virus load measured by both groups was only 10^8 – 10^9 particles, each with a half-life of about 1–2 days². This amount of virus would be produced by only 10^3 – 10^4 not 4×10^7 CD4⁺ cells per day. Alternatively, if the number of physical particles produced each day was actually 10^{12} – 10^{13} , then each released virion would have a half-life of approximately 10 s, which is unrealistic. Unfortunately, neither group directly measured the number of productively infected cells in their analyses of virus and CD4⁺ lymphocyte turnover during antiviral therapy. The number of newly infected cells per day was estimated from increases in CD4⁺ cell number, assuming, as noted above, a quasi-steady-state relationship between CD4⁺ lymphocyte replacement and killing.

Although the measured rise in circulating CD4⁺ cells is substantial, this increase may not be representative of total CD4⁺ cell turnover. Lymphocyte trafficking, homing and recirculation is a complex, multifactorial process and changes in the CD4⁺ cell number in the peripheral blood may not reflect simple quasi-steady-state alterations. Because lymphocytes in the peripheral blood comprise such a small fraction (1/50) of the total lymphocytes in the body, major changes can occur, for example, as a consequence of the rapid redistribution of cells from different compartments that attends normal diurnal variation⁸. Further, most lymphocytes do not circulate in either the blood or lymphatics but reside in spatially heterogeneous environments⁹. Thus the 50-fold multiplication step to calculate the total number of CD4⁺ cells killed by HIV-1 and replaced each day, which assumes rapid and complete exchange of lymphocytes between the peripheral blood and the rest of the body, is not warranted for such a large, heterogeneous non-equilibrated system unless the numbers and dynamic properties of cells residing in diverse major compartments are directly measured.

Dimitar S. Dimitrov

Laboratory of Mathematical Biology,
National Cancer Institute,

Malcolm A. Martin

Laboratory of Molecular Microbiology,
NIAID, National Institutes of Health,
Bethesda, Maryland 20892, USA

1. Ho, D.D. et al. *Nature* **373**, 117–122 (1995).
2. Wei, X. et al. *Nature* **373**, 117–122 (1995).
3. Dimitrov, D.S. et al. *J. Virol.* **67**, 2182–2190 (1993).
4. Graziosi, C. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6405–6409 (1993).
5. Clark, S.J. et al. *New Engl. J. Med.* **324**, 954–960 (1991).
6. Daar, E. S., Moudgil, T., Meyer, R. D. & Ho, D. D. *New Engl. J. Med.* **324**, 961–964 (1991).
7. Platak, M. et al. *Science* **259**, 1749–1754 (1993).
8. Olszewski, W. L. In *Vivo Migration of Immune Cells* (CRC, Boca Raton, 1987).
9. Young, A. J., Hay, J. B. & Mackay, C. R. *Curr. Topics Microbiol. Immun.* **184**, 161–173 (1993).

Antiviral therapy

SIR — Ho et al.¹ state that “patients with lower initial CD4 cell counts had more prominent rises” (in CD4 count during treatment). This conclusion was substantiated with an analysis showing a correlation coefficient of -0.57 ($P < 0.01$) for the association between initial CD4 count and slope of log CD4 count increase after therapy. This finding suggests there may be a greater propensity for CD4 lymphocyte regeneration at lower CD4 counts compared with higher counts. This could explain the paradox of why the CD4 count decline is on average more rapid early in HIV infection, when the CD4 count is high and cellular and cell-free virus are low, rather than later in the infection when CD4 counts are lower and virus levels are substantially higher².

However, although indeed the number of new cells appearing in the blood per existing cell is greater at lower CD4 counts, the absolute number of new cells appearing in the blood at any given time may not necessarily be greater, for which the slope of unlogged CD4 count increase after therapy must be measured. When this is done the correlation coefficient for the association between initial CD4 count and slope of CD4 count increase is 0.08 ($P = 0.75$; Spearman rank correlation coefficient), suggesting that on average no more new cells appear in the blood in a given time after starting therapy in individuals with lower CD4 counts than in those with higher counts. When looked at in this light, the results¹ do not appear to explain the paradox. The suggestion^{3,4} that the immune system is likely to play a major role in cell destruction — rather than the process being mainly due to direct cell killing by virus — is more consistent with the data as it offers a possible explanation for why the rate of CD4 count decline gradually decreases as more severe immunodeficiency develops.

Ho et al. also suggest¹ that analyses of the marked CD4 count rises seen after therapy “demonstrate convincingly that the CD4 lymphocyte depletion seen in AIDS is primarily a consequence of the destruction of these cells induced by HIV-1, not a lack of production”. Although this may be correct, it is not clear how the latter possibility has been excluded: if HIV-1 inhibited CD4 cell production, a rise in cell numbers might be expected when HIV-1 levels are drastically reduced by therapy. It is also not clear that the newly increased CD4 count in the blood is indeed due to newly pro-

duced cells. As the viral load decreases acutely, the redistribution of CD4 cells from tissues to blood, perhaps with an accompanied decrease in CD8 count⁵, is possible.

**Andrew N. Phillips, Caroline A. Sabin
Amanda Mocroft**

Department of Public Health,
George Janossy
Department of Clinical Immunology,
Royal Free Hospital School of Medicine,
London NW3 2PF, UK

Other approaches

SIR — Some^{1,2} have inferred from two recent papers^{3,4} that CD4⁺ T-cell infection determines parameters of HIV disease, and that immunostimulatory therapies are doomed to fail⁵. But alternative interpretations remain.

The new work^{3,4} reports that the turnover of CD4⁺ lymphocytes in AIDS patients is about 8 cells per μ l blood per day. Does this turnover reflect an accelerated rate of CD4⁺ T-cell production that strains the long-term proliferative capacity of the immune system? A definitive answer awaits accurate estimates of the turnover and half-life of both proliferating and peripheral CD4⁺ T cells in healthy individuals, normative data for which the immunological community strangely lacks a robust appraisal. However, taking 75 days (range of 50–100 days, as adopted in ref. 4) as a best current estimate of peripheral blood mononuclear cell half-life in the circulation, and $1,100$ CD4⁺ cells per μ l peripheral blood as an accepted average⁵, we can calculate that $0.5 \times 1,100/75$, or about 7, CD4⁺ T cells per μ l are normally replaced in the circulation each day. This turnover number is surprisingly close to that calculated in AIDS patients, suggesting that the infection-specific rate of CD4⁺ T-cell production is close to normal and unlikely to exhaust proliferating populations, even after years of disease. If so, immunostimulatory strategies should not yet be discounted as viable therapeutic approaches against AIDS.

A most remarkable observation by both groups^{3,4} is that, in each patient, post-treatment titres of resistant (mutant) viral strains rapidly rise very nearly to pretreatment levels. Why are the antiviral responses of the immune system, despite its activation against HIV-1, unable to maintain viral replication at lower levels once the infection has been reduced by chemotherapy? Although viral replication immediately after the emergence of resistant strains outstrips the antiviral defences of the immune system, the rate of increase of virus production eventually slows, establishing an equilibrium between production and clearance of the virus.

One simple hypothesis would account for this phenomenon. If CD4⁺ T-cell

1. Ho, D. D. et al. *Nature* **373**, 123–126 (1995).
2. Phillips, A. N. et al. *Nature* **367**, 124 (1994).
3. Wain-Hobson, S. *Nature* **373**, 102 (1995).
4. Zinkernagel, R. M. & Hengartner, H. *Immun. Today* **15**, 262–268 (1994).
5. Adelman, L. M. et al. *J. AIDS* **6**, 144–152 (1993).

infection depended, for instance, on cell-cell contact with a limited number of virus-presenting cells (for example, dendritic cells; ref. 6) in the lymphoreticular system, productive viral infections would begin to occur at a very high rate as drug-resistant strains spread through the population of antigen-presenting cells, each of which infects a large number of CD4⁺ T cells. Once most virus-presenting cells in a lymphoid organ are infected, the kinetics of virus production will dominantly reflect the rate of replacement of infected CD4⁺ T cells by uninfected cells at the loci of effective contact with virus-presenting cells, and will occur at a much slower rate. Under such a scheme, limiting the spread of HIV-1 infection within the virus-presenting cell population would represent a promising approach to the therapy of HIV disease⁷.

Michael Bukrinsky

Kirk Manogue

Anthony Cerami

The Picower Institute for Medical Research,

Manhasset, New York 11030, USA

1. Wain-Hobson, S. *Nature* **373**, 102 (1995).
2. Coffin, J. *Science* **267**, 483-5489 (1995).
3. Ho, D.D. et al. *Nature* **373**, 123-126 (1995).
4. Wei, X. et al. *Nature* **373**, 117-122 (1995).
5. Malone, J. L., Simms, T.E., Gray, G.C., Wagner, K.F. & Burge, J.R. *J. AIDS* **3**, 144-151 (1990).
6. Pope, M. et al. *Cell* **78**, 389-398 (1984).
7. Dubrovsky, L. et al. *Molec. Med.* **1**, 217-230 (1995).

Paradox remains

SIR — The articles by Ho *et al.*¹ and Wei *et al.*² have been hailed as providing crucial new information that clarifies the enigma of HIV-mediated pathogenesis (see, for example, refs 3, 4). To the extent that they have estimated equilibrium rate constants and provided an explanation for rapid development of drug resistance, these studies^{1,2} do provide new and important information. But the central paradox of AIDS pathogenesis remains.

The new studies on the dynamics of HIV infection demonstrate that the underlying rate of virus production is still 4-20 times lower than the rate of cell turnover. Because each infected cell produces many virions, and a high proportion of them are known to be defective, there is about 100-1,000-fold more cell death than can be accounted for by the observed rate of virus production⁵. It is a murder scene with far more bodies than bullets.

Instead of addressing this discrepancy, the authors^{1,2} suggest that "virus must be underestimated", or that many more infected cells must be hiding in deep lymphoid organs where they cannot be observed. Taken at face value, however, these data demonstrate that the CD4 lymphocyte population is highly activated, and that the fate of uninfected cells may be much more important than that of infected cells in AIDS pathogenesis. This

view is consistent with observations that the rate of spontaneous apoptosis in the peripheral lymphocytes of HIV-infected individuals is both sharply elevated and at least 10-100-fold higher than the frequency of productively infected cells⁶.

Ho *et al.* suggest the analogy of a sink, with the tap and drain both equally wide open, which eventually empties because the "regenerative capacity of the immune system (the tap) is not infinite" and cannot quite keep up. However, the differential between the tap and drain is extremely small (20-200 × 10⁶ cells per day) compared with the overall flow rate (2 × 10⁹ cells per day), and must remain relatively fixed for an average of 10 years, as CD4⁺ cell loss is roughly linear throughout most of the natural history of HIV infection⁷. Further, this model would predict no CD4 loss until virus production exceeded a critical threshold, and then an accelerating cell loss as the virus burden increases.

A more plausible explanation for these data is that a mechanism that finely regulates CD4 replacement makes a slight error, resulting in the failure to completely replace the cells (infected and uninfected alike) which are lost through programmed cell death, the natural consequence of immune activation. We have argued that this is exactly what would be expected if the immune system were exposed to a persistent co-stimulatory T-cell signal caused by the interaction of gp120 with CD4 (ref. 8). This extra signal causes the control mechanism to sense an elevated state of immune responsiveness and downregulate the recovery of "memory" cells to prevent growth of the immune system⁹. The result would be an inexorable but steady decay based on this difference in probability of survival.

Maddox⁴ wonders "why, after more than a decade of research, has it only now emerged that the response of the immune system to infection by HIV is hyperactivity rather than the opposite?". The answer to this question is that those who would see AIDS as a more-or-less conventional viral infection have consistently refused to recognize the paradoxes that are clearly evident in the experimental data. The problem continues.

**Michael S. Ascher, Haynes W. Sheppard
Russell W. Anderson, John F. Krowka
Hans J. Bremermann**

*Viral and Rickettsial Disease Laboratory,
California Department of Health Services,
Berkeley, California 94704, USA*

1. Ho, D. D. et al. *Nature* **373**, 123-126 (1995).
2. Wei, X. et al. *Nature* **373**, 117-122 (1995).
3. Wain-Hobson, S. *Nature* **373**, 102 (1995).
4. Maddox, J. *Nature* **373**, 189 (1995).
5. Sheppard, H. W. et al. *Nature* **364**, 291 (1993).
6. Gougeon, M. L. & Montagnier, L. *Science* **260**, 1269-1270 (1993).
7. Sheppard, H. W. et al. *J. AIDS* **7**, 1159-1166 (1993).
8. Ascher, M.S. & Sheppard, H. W. *J. AIDS* **3**, 177-191 (1990).
9. Sheppard, H. W. & Ascher, M. S. *J. AIDS* **5**, 143-147 (1992).

SIR — D. Ho *et al.* (*Nature* **373**, 123-126; 1995) state that HIV decay slopes (clearance rate constants) are independent of the initial viral loads, and that the slopes do not correlate with the initial CD4 lymphocyte counts. They conclude that the viral decay slope is not dependent on the stage of HIV-1 infection (their Fig. 1b).

But regression analysis of the same data shows that the correlation between HIV RNA (copies per ml) and the viral decay slopes is 0.17, with a *P* value of 0.47. When we stratify study subjects' viral decay rates according to levels of HIV RNA with a cut-off point of 100 (<100 versus ≥100), however, the viral decay rate in the stratum with lower RNA is significantly different from that in the stratum with higher RNA (Kruskal-Wallis test: *P* = 0.048). The median of the viral decay rate in the lower RNA stratum was -0.28, while the median in the higher RNA stratum was -0.38. A univariate linear regression analysis shows that the dichotomous variable RNA (<100 versus ≥100) was significantly associated with the viral decay rate (*P* = 0.026). These analyses suggest that when RNA copies were at higher levels, the viral decay slopes would be higher. In other words, the viral decay slope is a function of many factors, including initial viral load and the stage of HIV infection.

Ho *et al.* also state that the slopes are inversely correlated with the baseline CD4 counts. Exponential modelling of CD4 increases reveals that the slopes are inversely correlated with the initial CD4 lymphocyte counts; the authors conclude that the CD4 lymphocyte depletion seen in AIDS is primarily a consequence of the destruction of these cells induced by HIV-1, not a lack of their production. The increase in CD4 lymphocyte counts following ABT-538 administration was also modelled linearly and the authors claimed that although their two sets of analyses (modelling both exponential and linear increases) do not yield identical numerical results, they are in close agreement and emphasize the same qualitative points about HIV-1 pathogenesis.

Based on our univariate linear regression analysis, the slope of increase generated by an exponential growth model is significantly and inversely associated with the baseline CD4 counts (*P* = 0.01). Nevertheless, the slope of increase generated by a linear production model was not significantly correlated with the baseline CD4 (*P* = 0.17). After adjusting for the viral loads, measured by HIV RNA, both the linear and exponential slopes seem to be correlated with the baseline CD4 counts, but in opposite directions. With the slope obtained from a linear production model, the slope was, although borderline, significantly and positively associated with the baseline CD4, after controlling RNA levels (*P* = 0.06). This implies that the slope

generated from a linear model could be a function of both the baseline CD4 level and viral loads. Further studies with larger sample sizes are needed to resolve these discrepancies.

Shenghan Lai, J. Bryan Page, Hong Lai
Departments of Medicine,
Psychiatry and Epidemiology,
University of Miami School of Medicine,
Miami, Florida 33136, USA

HIV an illusion

SIR — In an editorial¹ in the 19 January issue of *Nature*, John Maddox invited “Duesberg and his associates” to comment on the “HIV-1 dynamics” papers published the previous week, indicating that these new results should prove an embarrassment to us. Although we do not think that a scientist should be embarrassed for pointing out inconsistencies and paradoxes in a hypothesis that have only been reportedly resolved 10 years later, we nonetheless prepared a fully referenced, approximately 2,000-word critique of the Ho *et al.*² and Wei *et al.*³ papers that we believed met the criteria of “not being longer than it needs to be, and pertaining to the papers at hand” that Maddox set out in his widely read challenge.

Unfortunately, he did not share our view and agreed to publish only a radically shortened version, and only after he had personally “gone over it with a fine-tooth comb” to remove our perceived misrepresentations of the issues. We found these new conditions so totally at variance with the spirit of free and fair scientific debate that we could not agree to them.

Readers of *Nature* who are interested in these questions, and feel that they do not need to be protected by Maddox from our ill-conceived logic, can find the complete text of our commentary in the monograph supplement to the most recent issue of *Genetica*⁴. Here we would point out only that the central claim of the Ho *et al.*² and Wei *et al.*³ papers — that 10^5 HIV virions per ml plasma can be detected in AIDS patients with various nucleic-acid amplification assays — is misleading. The senior author of the Wei *et al.* paper has previously claimed that the PCR method they used overestimates by at least 60,000 times the real titre of infectious HIV⁵: $100,000/60,000$ is 1.7 infectious HIVs per ml, hardly the “virological mayhem” alluded to by Wain-Hobson⁶. Further, Ho and a different group of collaborators have just shown⁷ that more than 10,000 “plasma virions”, detected by the branched-DNA amplification assay used in their *Nature* paper, correspond to less than one (!) infectious virus per ml. And infectious units, after all, are the only clinically relevant criteria for a viral pathogen.

Finally, in view of Wain-Hobson’s statement⁶ that “the concordance of their [Wei

and Ho’s] data is remarkable”, note that Loveday *et al.*⁸ report the use of a PCR-based assay and find only 200 HIV “viral RNAs” per ml of serum of AIDS patients — 1,000 times less than Ho and Wei. So much for the “remarkable concordance”.

Peter Duesberg

Department of Molecular and Cellular
Biology, University of California,
Berkeley, California 94720, USA

Harvey Bialy

Bio/Technology, New York,
New York 10010, USA

1. Maddox, J. *Nature* **373**, 189 (1995).
2. Ho, D.D. *et al.* *Nature* **373**, 123–126 (1995).
3. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).
4. Duesberg, P. & Bialy, H. *Genetica* Suppl. (in the press).
5. Piatak, M. *et al.* *Science* **259**, 1749–1754 (1993).
6. Wain-Hobson, S. *Nature* **373**, 102 (1995).
7. Cao, Y. *et al.* *New Engl. J. Med.* **332**, 201–208 (1995).
8. Loveday, C. *et al.* *Lancet* **345**, 820–824 (1995).

■ Peter Duesberg was offered space in Scientific Correspondence for 500 words of his own choice, but declined. — Editor, Scientific Correspondence.

SIR — The most recent AIDS press releases published as two “scientific articles” in your magazine^{1,2} indeed are the most clever and, if possible, the most deadly yet! Since March 1987, when Duesberg first published his critique of retrovirology³ and his proof that the “AIDS virus” is not sufficient to cause AIDS, the AIDS/HIV entrepreneurial gaggle have railed against him both professionally and financially. Rather than test Duesberg’s hypothesis or for that matter fund him to test his own hypothesis, you change your model, which has not saved a single life, to suit your own purposes, from HIV kills CD4 cells directly, to HIV kills cells indirectly, and now back to the latest direct kill model of Wei *et al.*¹ and Ho *et al.*².

Wei *et al.*¹ and Ho *et al.*² describe a viraemia that on administration of certain drugs drops exponentially and raises CD4 counts exponentially, seemingly without a trace of “mayhem”. What is this viraemia of billions of RNA particles that can only be seen with an undocumented branch-PCR or PCR but not with a functional infectivity test? How do the authors know that these RNA particles are infectious and not defective? Is it reverse transcriptase activity? Is it some protein that is unique to the quasispecies HIV? The tests for HIV are all flawed. What is this free virus? How does one have free virus after immune response? Are not symptomatic people with AIDS HIV-positive, that is, don’t they have antibodies against HIV? So what is free virus?

The Wei *et al.* and Ho *et al.* models do not account for other probable immune activity and the uniqueness of each individual. Haynes and Fauci showed in 1978 that hydrocortisone selectively caused CD4 cells to hide in tissue. Indeed, what effect do multiple drugs have on CD4 cells? The notion of clearance of one billion CD4 cells per day versus one billion

virions per day is not defined or addressed, nor can it be! No one else has access to either the unapproved drugs or the branch PCR technology! What is the benchmark rate for the turnover of CD4 cells in the general population?

To counter the 42 case studies of Wei *et al.*¹ and Ho *et al.*², we at HEAL (Health Education AIDS Liason) can provide at least 42 people who are western-blot-positive for ‘HIV’, have low T4 cells, who are not using orthodox procedures, and have been healthy for years! On the other hand, we can also provide you with hundreds of ‘HIV⁺’ people with high T4 cells, who were then hospitalized with opportunistic diseases! Not to mention the thousands of AIDS cases without HIV which were conveniently renamed idiopathic CD lymphocytopenia.

We at HEAL maintain that what is called AIDS is no more complicated than a recreational drug-fear-medical drug disease syndrome which, with the exception of acute medical care, is out of the purview of orthodox medicine.

F. R. Buianouckas

Department of Mathematics
and Computer Science,
Bronx Community College,
City University, Bronx,
New York 104353, USA

1. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).
2. Ho, D.D. *et al.* *Nature* **373**, 123–126 (1995).
3. Duesberg, P. H. *Cancer Res.* **47**, 1199–1220 (1987).

Toxic shock

SIR — The recovery rate of CD4⁺ T cells reported by Wei *et al.*¹ and Ho *et al.*² in HIV patients following therapy is remarkably similar to that observed in another acute clinical condition. We have documented lymphocyte counts and proportions of CD4 cells double-staining for CD45 isoforms in three patients during episodes of toxic shock syndrome and one following toxic shock-like syndrome.

In these clinically defined conditions, bacterial superantigen initially induces massive T-cell and monocyte activation. There is subsequent apoptosis of the activated CD4⁺ cells, during which patients demonstrate a lymphopenia³, followed by the return to normality of the lymphocyte count as patients recuperate. The mean rate of change in the counts in the first 2 weeks of recovery was 2.1, which compares with 2.0 (ref. 1) and 1.8 (ref. 2).

The similarity in rates of recovery suggest that this rate has some value as a common measure of maximal CD4 T-cell output to the circulation in adults. Slower rates may be observed, for instance, in patients following radiotherapy or chemotherapy⁴. This rate decreases with time; it took one patient 14 months to reach a final CD4⁺ lymphocyte count of 1.4×10^9 cells cm⁻³. The nature or mechanisms of the brake on this rapid prolifer-

ation is not clear. Recovery in CD4⁺CD45RA⁺ numbers is the consequence of thymic output (probably small in adults⁴), proliferation of CD45RA⁺ T cells, together with reversion of primed CD45RO⁺ T cells⁵. Reversion rates have been estimated, and, if correct, are unlikely to contribute greatly to the recovery of the CD45RA⁺ lymphocyte count within the short period observed⁶. Proliferation of CD4⁺CD45RA⁺ cells is therefore likely to be the principal component involved in the recovery phase of the immune system following a large insult.

Colin Michie

Department of Paediatrics,
Ealing Hospital NHS Trust,
Middlesex UB1 3HW, UK

1. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).
2. Ho, D.D. *et al.* *Nature* **373**, 123–126 (1995).
3. Michie, C.A. *et al.* *Clin. exp. Immun.* **98**, 140 (1994).
4. Mackall, C.L. *et al.* *New Engl. J. Med.* **332**, 143 (1995).
5. Michie, C.A., McLean, A.R., Alcock, C. & Beverley, P.C.L. *Nature* **360**, 264–265 (1992).
6. McLean, A.R. & Michie, C.A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Cyclosporin A

SIR — Ho *et al.*¹ and Wei *et al.*² draw attention to the dynamics of HIV replication *in vivo*, through close observation of the response of plasma viraemia, quantified by RNA PCR, to antiretroviral therapy. In all subjects, HIV RNA levels fell rapidly over the first 7 days of therapy at a constant rate. Reduction in viral load was associated with a prompt elevation in CD4 number, which appeared to be related to interruption of viral replication consequent to antiretroviral therapy. The rapid reduction of HIV viral load by RNA PCR and parallel CD4 increase has been observed with diverse drugs³.

We demonstrate a novel, distinct pattern of reduction in viral load in HIV infection that is slow, sustained and associated with a parallel reduction in CD4 number. We believe this observation has implications for the pathogenesis of HIV infection.

The patient, a 36-year-old HIV-1-infected homosexual male, developed psoriasis in 1991; diverse topical and systemic agents were prescribed, without clinical benefit. By June 1993, the patient was incapacitated by psoriatic arthropathy and generalized pustular psoriasis, and 250 mg per day cyclosporin A (CyA) was given as a last resort, in the full knowledge that HIV infection may be accelerated by concomitant immunosuppression. There was a dramatic clinical response: there was a reduction of viral load of 84% ($0.87 \times \log_{10}$) over 30 weeks, which was sustained. But in contrast to the data presented in refs 1 and 2, CD4 number fell from 960 to $530 \times 10^6 \text{ l}^{-1}$ over this period. The relationship of these changes to CyA therapy was demonstrated when the patient ceased therapy about week 50–54, on account of an unrelated

alcoholic hepatitis. Viral load and CD4 rose to baseline off therapy, and fell again when CyA was recommenced on week 70. No other therapy was given over this entire period.

CyA is known to inhibit T-cell activation; we believe that these data provide evidence *in vivo* for the importance of T-cell activation in permitting HIV replication. CyA reduces CD4⁺ T-cell number, but also inhibits HIV replication by removing the host cell target. There has been speculation that there seems to be a ceiling *in vivo* for the absolute level of plasma viraemia; our observations suggest that the level of plasma viraemia may be determined by the level of availability of the activated target CD4⁺ cell. The use of glucocorticoids to reduce T-cell activation has recently been reported, but although these lead to an increase in CD4 cells, viral load is unaltered⁴.

Jonathan Weber, Sarah Galpin

Department of Genito-Urinary Medicine &
Communicable Diseases,
St Mary's Hospital Medical School,
London W2 1NY, UK

1. Ho, D. *et al.* *Nature* **373**, 123–126 (1995).
2. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).
3. Coffin, J. *Science* **267**, 483–489 (1995).
4. Andrieu, I. *et al.* *J. Infect. Dis.* **171**, 523–527 (1995).

HO *ET AL.* REPLY — The most substantive point raised by the letters printed above concerns CD4 lymphocyte redistribution, rather than proliferation, as an explanation for the rise in CD4 cell counts following treatment with potent antiretroviral agents. Although lymphocyte re-trafficking is a plausible explanation, as pointed out by Mosier, Sprent and Tough, Dimitrov and Martin, and Phillips *et al.*, several observations collectively argue against this suggestion.

First, the elevations in CD4 cell counts were not transient, but were sustained as long as the antiviral effect was maintained, in some cases for more than 6 months. Second, in other studies, CD4 lymphocyte increases associated with viral suppression were accompanied by significant clinical improvement¹. Third, in our recent unpublished studies, the surface-marker phenotypes of CD4 lymphocytes post-therapy differ substantially from those before treatment. In particular, our preliminary data reveal the expression of a number of activation markers on many of the CD4 lymphocytes after treatment, a finding that supports lymphocyte repopulation by cellular proliferation but argues against lymphocyte redistribution, because, as stated by Sprent and Tough, “most of the CD4⁺ cells would be expected to display a resting rather than an activated phenotype” if the latter were true.

In reanalysing our data, Mosier found a positive correlation between the exponential slope of the CD4 lymphocyte increase and the baseline plasma level or the viral turnover rate, whereas we showed an

inverse correlation with the baseline CD4 lymphocyte count². These findings are fundamentally the same, as all three parameters (viral load, viral turnover rate and CD4 count) reflect the baseline disease status of the patient; they do not necessarily support the lymphocyte redistribution hypothesis. To address properly his notion that antiviral treatment liberates CD4 lymphocytes from trapping in lymphoid tissues, Mosier should instead correlate the slope (perhaps the linear slope would be more appropriate) of the CD4 lymphocyte increase with the magnitude of the antiviral effect.

Bukrinsky *et al.* are incorrect to suggest that we^{1,2}, Wain-Hobson³ or Coffin⁴ have stated or implied “that immunostimulatory therapies are doomed to fail”. In addition, we do not understand their logic of comparing our calculated CD4 lymphocyte turnover rates with previous estimates for normal peripheral blood mononuclear cells, which are a mixture of B cells, CD8 lymphocytes, monocytes and divergent populations of CD4 lymphocytes.

Dimitrov and Martin question our estimates of the CD4 lymphocyte turnover rate based on a mathematical calculation. They suggest that if each infected lymphocyte produces 10^5 virions⁵, then the particle turnover rates are simply too low, perhaps by four orders of magnitude, given our estimated number for infected cells. However, in their derivations, they did not take into consideration that many infected lymphocytes *in vivo* may be eliminated by immune responses before release of progeny virions. In essence, the “bullets” and “bodies” analogy of Ascher *et al.* suffers from the same point. Furthermore, in a recent follow-up study to determine more accurately the virion turnover rates, we have found that there are indeed more “bullets” than “bodies”.

Phillips *et al.* and Lai *et al.* correctly point out that the linear (rather than exponential) slope of the CD4 lymphocyte increase does not correlate with the baseline CD4 cell count. However, the use of the linear model is based on the assumption that the CD4 lymphocytes are coming from a source such as the thymus. Based on lymphocyte phenotype studies, we now have evidence that the increase in CD4 cell counts is largely due to post-thymic lymphocyte proliferation.

David D. Ho

Aaron Diamond AIDS Research Center,
New York, New York 10016, USA

Alan S. Perelson

Los Alamos National Laboratory,
Los Alamos, New Mexico 87545, USA

George M. Shaw

University of Alabama at Birmingham,
Birmingham, Alabama 35294, USA

1. Ho, D. D. *et al.* *Nature* **373**, 123–126 (1995).
2. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).
3. Wain-Hobson, S. *Nature* **373**, 102 (1995).
4. Coffin, J. *Science* **267**, 483–489 (1995).
5. Dimitrov, D. *et al.* *J. Virol.* **67**, 2182–2190 (1993).

For science and country

David Cahan

Fritz Haber: Chemiker, Nobelpreisträger, Deutscher, Jude. By Dietrich Stoltzenberg. VCH: 1994. Pp. 669. DM98.

THE German physical chemist Fritz Haber (1868–1934) was a gifted scientist and science organizer. He was also, thanks to his leadership in the development and use of chemical weapons during the First World War, one of the most controversial and reviled scientific figures of his age. His aggressive German patriotism made him into a villain for many in England, France and elsewhere, while his Jewish ancestry later turned him into an object of Nazi persecution. His life and work after 1914 were filled with drama and painful contradictions. Despite several scholarly articles and one old short volume that leaves much to be desired, we have long needed a substantial biography of him.

Dietrich Stoltzenberg, a retired German chemist, has now fulfilled this need. His exhaustive book draws not only on the published scientific papers but also on many previously unknown archival documents relating to Haber. Although he sometimes provides too much quotation and minutiae, as well as making the occasional whiggish reference to more recent scientific and political developments, the author generally does an excellent job in examining Haber's scientific achievements and career, revealing the ups and downs of his life and, if only implicitly, pointing out the flaws in his character.

Stoltzenberg provides a level of detail about Haber's private life that few scientific biographies have ever matched. We learn a great deal about Haber's family and childhood in Breslau (now Wrocław in Poland). It is an important and fateful fact that he was born of partially assimilated Jewish parents (his father was a merchant) who barely practised any Jewish rites and seemed indifferent to matters of belief, yet whose lives were deeply enmeshed in the Breslau Jewish community. In 1892, at the age of 24, Haber renounced Judaism and was baptised into the Evangelical faith, a faith he never practised — he apparently converted simply to aid his pursuit of an academic career. Stoltzenberg portrays Haber as a man of great charm but as someone who, until 1933, seems to have been lacking in a spiritual dimension and a feeling for the needs of others. Apart from pursuing his scientific work, Haber wanted little more in life than to belong to the German military, political and industrial élite. He dedi-

cated himself entirely to his work and to what he considered to be his country's interests.

Stoltzenberg tells us about Haber's scientific education at Berlin, Jena and Karlsruhe. In 1899, Haber graduated at Berlin with an undistinguished dissertation and oral examination in organic chemistry. Inspired by influential physical chemists such as Wilhelm Ostwald, he decided to study physical chemistry at the Karlsruhe Polytechnic. Although Haber

was largely an autodidact, the polytechnic did teach him the importance of applying his chemistry practically. While working to qualify as an assistant in physical chemistry and then ascending the academic ladder to a professorship in 1906, Haber forged many contacts in the German chemical (especially electrochemical and dye) industries, not least the nearby Badische Anilin und Soda-Fabrik (BASF).

His time at Karlsruhe (1894–1911) proved to be his most creative scientifically. His many accomplishments included the first general theoretical studies of electrochemical reduction and studies on the electrochemistry of self-oxidation and on chemical thermodynamics. What is more, during a trip to the United States in 1902, he learnt about a US chemical company's attempt to 'fix' nitrogen from the atmosphere. In a long chapter, Stoltzenberg traces Haber's route to his discovery of a process for nitrogen fixation (the synthesis of ammonia from free nitrogen and oxygen), his best known — and perhaps

most important — work in chemistry. This work became intimately connected to industrial needs, being developed for large-scale use by the BASF chemist Carl Bosch around 1913. During the First World War, the Haber–Bosch nitrogen-fixation process proved to be essential in helping Germany to meet its agricultural and military needs. Haber was awarded the 1918 Nobel prize in chemistry for his discovery.

Haber built up and inspired loyal, productive and creative research groups, both at Karlsruhe and at Berlin, where he moved in 1911 to be the first director of the first institute of the Kaiser Wilhelm-Gesellschaft (KWG), an organization consisting of institutes devoted solely to advanced scientific research. As Stoltzenberg nicely shows, Haber became one of the leading figures in the development of the entire KWG, which after the Second World War was renamed the Max Planck-Gesellschaft. During his two decades as director of the Institute for Physical Chemistry and Electrochemistry, Haber helped to establish Berlin, if not his own institute, as the world's centre for physical chemistry. In 1920, he tried to protect German science from the effects of runaway inflation by becoming one of the founders of the *Notgemeinschaft der deutschen Wissenschaft*, the forerunner of modern Germany's *Deutsche Forschungsgemeinschaft*.

Haber paid a high price for his scientific and technical successes at Karlsruhe and Berlin. For one, he neglected his two wives (the first committed suicide and the second divorced him) as well as his three children (one of whom also committed suicide). On a different note, he exploited his institute, ironically founded *inter alia* to advance peace, as a platform from which he infamously initiated and directed Germany's development and use of chemical weapons during the First World War. He believed that his work was patriotic and that these weapons — he often went to the front lines to oversee their implementation and assess their effects — were no worse and perhaps more humane than other types of warfare. The Allies did not think so: in 1920, they listed Haber as a war criminal. Perhaps their initiative was partisan; yet, as Stoltzenberg goes on to show, after the First World War, Haber, aided in good part by Stoltzenberg's own father, the chemist Hugo Stoltzenberg, secretly led the German military's attempts to develop chemical weapons further and to supply them to the new Soviet government. The unceasing military effort during the early Weimar period calls into question Haber's moral justification of his military work.

Mary Evans

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Fritz Haber: a "deeply flawed individual".

All this led to ruined health and spiritual impoverishment. Haber's personal and political travails, described by Stoltzenberg in a surfeit of detail, reached their climax when the Nazis came to power in 1933. Despite Haber's patriotic war efforts and his scientific leadership, the Nazis drove him into exile (he worked for a while at the Cavendish Laboratory in Cambridge, England) and finally to his death. They unwittingly forced him to confront his Jewish heritage and his treatment of his family: he seriously considered moving to Palestine and hoped that his sons would become Englishmen. He died in Basel in January 1934, a broken, confused, lonely and bitter man.

Yet the ultimate contradiction and tragedy of Haber's life and work was yet to come: during the Second World War, the Nazis used the gas Zyklon B, studies of which were first done in Haber's own institute around 1920, in Auschwitz and other concentration camps. Members of Haber's own family were among the vast numbers of people murdered with this poison.

Stoltzenberg has written a fine biography of this deeply flawed individual. Graced as it is by nearly a hundred excel-

lent photographs, it will appeal to a wide variety of readers and deserves to be translated into English. To be sure, only those with a basic knowledge of physical chemistry will be able to take the full measure of Stoltzenberg's treatment of Haber's scientific work. But the sympathetic and comprehensive account of Haber's personal life and friendships (especially with such important personalities as Albert Einstein and Richard Willstätter), his organizational activities in science, his efforts to develop industrial and political support for German science in general, his development of gas warfare and his application of chemistry in general, his receipt of the Nobel prize, his involvement with many of Germany's scientific, industrial and political leaders, and his Jewish ancestry and ambivalence towards the Zionist cause, should appeal to general readers as well as to historians and all those interested in the social responsibility of science. □

David Cahan is in the Department of History, University of Nebraska, 612 Oldfather, PO Box 880327, Lincoln, Nebraska 68588-1327, USA.

Organizing entomological miracles

Tristram D. Wyatt

Insect Hormones. By H. Frederik Nijhout. Princeton University Press: 1994. Pp. 267. \$35, £27.50.

INSECT metamorphosis is one of the most spectacular phenomena of the living world, but controlling it is only one of the many roles for hormones in the biology of insects: almost all insect activities are mediated by hormones, from the control of reproduction to internal homeostasis, through polyphenism (for example castes in termites) and diapause, to behaviour in spinning a cocoon. As model biological systems, including investigations of animal development, receptors and second messengers, insect hormones and their targets have an important influence far outside entomology.

The study of insect hormones offers many classic examples of the power of choosing the 'right animal'. To study diuretic hormones, for example, use the blood-sucking bug *Rhodnius prolixus*, which can excrete urine in quantities of up to a prodigious six times its body weight in the hour after a blood meal. Unfortunately, investigating just a few species has disadvantages. Insects show great variations in the hormone molecules involved in many processes. The patterns offer a fascinating paradox. Not only are there the unifying characteristics revealed by new molecular approaches showing similarities between some verte-

brate hormone sequences and insect neuropeptides (for example bombyxin and vertebrate insulin-like growth factors); there is also enormous adaptive radiation among insects, with different hormones used for the same function, or the same hormone used for quite different ends in different species or at different life-stages by the same insect. The hormone-receptor-intracellular signalling system offers great scope for flexibility in evolution.

It is this complex situation that Nijhout tackles well: he gives a clear outline and helpful synthesis. He emphasizes areas where answers are still unknown or unclear, without leaving the reader at sea (although a few more diagrams would help). The book successfully offers students and researchers (including those in other fields) a lucid overview, although some may wish for more detail here and there; half the text is devoted to metamorphosis and development, for example, whereas caste determination in ants is covered in just four pages.

Molecular investigation of hormones is becoming ever more sophisticated and productive. While covering these recent advances, this timely book returns the focus to the way in which these molecules work in the insect as a whole organism. □

Tristram D. Wyatt is in the Departments of Continuing Education and Zoology, University of Oxford, Oxford OX1 2JA, UK.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 21 September. Publishers and learned societies are invited to submit journals for review, taking note of the following criteria:

■ Journals that first appeared during or after June 1993 and issued at least four separate numbers by the end of April 1995 will be considered.

■ Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are publications of abstracts.

■ Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are, of course, eligible.

■ Deadline for submission is the end of May.

When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates (personal and institutional) and frequency of publication to: Peter Tallack, *Nature*, Macmillan Magazines Ltd, Porters South, Crinan Street, London N1 9SQ, UK. Tel: +44 (0)171 843 4567.

New in paperback

■ Oxford University Press has just reissued paperback editions of James Lovelock's influential books on his Gaia theory, which views the Earth as a self-regulating system or 'superorganism'. Lovelock first sketched out his theory 15 years ago in *Gaia: A New Look at Life on Earth* (£6.99). He amplified his ideas several years later in *The Ages of Gaia: A Biography of Our Living Earth* (£7.99; also published in the United States by Norton at \$12). He has now fully updated this volume and reluctantly revised it to make it "scientifically correct". In vivid and often poetic prose, he assures us that it is in no way the guide book of a New Age cult. He also rails against the narrowness and dogma that he believes have prevented modern scientists from accepting Gaia.

■ **Abduction: Human Encounters With Aliens** by John E. Mack. Simon and Schuster, £5.99. In this bestselling book, a professor of psychiatry at Harvard Medical School presents his 13 case studies of "authentic and disturbing mysteries" that he claims cannot be dismissed simply as psychological phenomena. His work is now rumoured to be at the centre of a special investigation by a faculty committee set up by Harvard University (see *Nature* **375**, 5; 1995).

Life story

Andrew H. Knoll

Vital Dust: Life as a Cosmic Imperative. By Christian de Duve. *Basic Books*: 1995. Pp. 362. \$25, £16.99.

On my first trip to London, 25 years ago, I chanced to see Ingrid Bergman in a West End production. She flubbed numerous lines; I was enchanted. It would have been hard not to be won over, because Bergman radiated an unmistakable delight in the encounter with play and audience, and joy in the maturity of a life spent in theatre. That evening came to mind as I read Christian de Duve's monumental attempt to understand life's origin, evolution, future and meaning. De Duve's all-encompassing enquiry provides numerous opportunities for disciplinary quibbling, but his sweep of vision and joy in a lifelong encounter with biology overshadow the occasional flubbed line.

The first two-thirds of *Vital Dust* is an explanation of how we got here. 'We' is meant literally, for de Duve is not one to view *Homo sapiens* as a small twig on the tree of life. We are paramount, the logical culmination, if not the objective, of evolution. Indeed, in one of the book's few illustrations, the tree of life is shown resplendent in the dense foliage of species diversity and then stripped of obfuscating leaves to reveal a trunk that rises vertically from its progenetic base to an apex marked "humans".

Revisiting arguments advanced in his earlier book, *Blueprint for a Cell*, de Duve reasons that clues to life's origin must be sought primarily in the cell rather than in test tubes. Prebiotic syntheses are of little significance unless they are "congruent", to use de Duve's term, with biochemistry. The need for excess hydrogen, the problems of condensation in an aqueous milieu, and the requirement for energy-transfer collectively lead him to identify thiols and thioesters as key elements in a primordial metabolism that arose amid the iron- and sulphur-rich environments of hydrothermal vents. ATP emerged early and, de Duve admits, mysteriously as the energy currency for nascent biochemistry. Poly(A) oligomers arose as a means of storing AMP, but later evolved informational complexity, ushering in the RNA world. It's a good story, but the difficulty of abiotic nucleotide synthesis constitutes a major stumbling-block.

Throughout de Duve's provocative speculations, two themes surface repeatedly. The origin of life is not the chance product of unlikely reactions; it is the expected consequence of chemistry in prebiotic environments on Earth and, presumably, elsewhere — de Duve's cos-



SCIENTIFIC blueprints — the architect Charles Jencks is well known as the author of many influential books on contemporary building and postmodern thought. In *The Architecture of the Jumping Universe* he now blends ideas from art, design, science and philosophy in a polemical exploration of how cosmology and complexity theory are changing architecture and culture. To most scientists the pictures will speak louder than the words. Shown here are Jencks's "Universe Model" and "Soliton Gate". Published in paperback by Academy Editions, £12.95 (distributed in the United States by St Martin's Press).

mic imperative. Chance plays an equally small role in the progression from primordial metabolism to complex biochemistry, or, for that matter, from bacterium to human; natural selection determines the vector of evolution.

Vital Dust may be the first account of life's history written expressly, and almost exclusively, from the perspective of cell biology. De Duve's fossil record is the genome, and he is not much at home with the more conventional fossils preserved in rocks. Indeed, in his introductory chapter, de Duve asserts that "the whole history of life is written into present-day organisms". If so, it must be a history devoid of trilobites, dinosaurs and countless other extinct organisms that convey a central evolutionary message not encrypted in our genes. Evolution did not just produce the life currently resident on Earth; it gave rise to a far greater diversity of organisms *different* from those now alive. De Duve gives particularly short shrift to the role of environmental dynamics in influencing evolution, but in the one paragraph devoted to this subject, a single magisterial sentence captures the significance of environmental history: "Apparently, when evolution becomes sluggish, it is not so much for want of an appropriate chance mutation as for the lack of a worthy environmental challenge."

The evolutionary contingency implied by this statement is not, however, central to de Duve's view of life. Metabolic complexity, multicellularity, the nervous sys-

tem, intelligence — all are more or less expected in due course once life is established. Mind you, there is not a whiff of vitalism here; de Duve's view is that the determinate nature of chemistry and the strong selective advantage conferred by certain mutations make the principal features of evolutionary history unavoidable. Readers must decide for themselves whether the dichotomy between this biological determinism and the contingency of environmental disruption and multiple adaptive peaks can be resolved in blind-men-and-elephant fashion.

In the final third of *Cosmic Dust*, de Duve's profoundly deterministic viewpoint provides his answers to the great philosophical questions of life. Consciousness seems to have arisen at the moment when neuronal circuits first fired in response to a hypothetical rather than physical stimulus, but, although born of cellular function, it transcends structural chemistry because it endows humans with the ability to search for meaning in the evolution of DNA and proton pumps, of homeoboxes and axons. For de Duve, it is the deterministic chemistry of life and the concomitant probability that the Universe is full of biospheres much like our own that gives philosophical meaning (but not necessarily purpose) to our existence. But, if we accept the alternative view that human intelligence is unlikely and perhaps even unique, does that make the Universe any more or less meaningful? Readers will leave this section with many different

opinions, fashioned, I suspect, by neuronal networks wired by individual scientific and social experience.

In the end, it is not de Duve's answers that make *Vital Dust* important. Its significance — and the sheer enjoyment it provides — lies in the struggle with great questions and the exuberant drive to explain the intricate complexities of living things. Anyone who contemplates biology seriously will eventually be called to a similar journey. What good fortune that we have de Duve's elegant volume for our Baedeker. □

Andrew H. Knoll is in the Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

All consuming

David Moore

The Physiology of Fungal Nutrition. By D. H. Jennings. Cambridge University Press: 1995. Pp. 622. £85, \$150.

In these days of research assessment exercises that denigrate scholarship as not contributing to the research quality of a department, it may be ungenerous to describe a book as a work of scholarship. However, and I mean this in the best possible sense and with sincerity, here is a work of high scholarship that will serve as a source of guidance and reference for many years to come.

The treatment is comprehensive (there are 126 pages of references) and up to date, but it is also rigorous and critical. Dubious experimental approaches and over-indulgent claims do not escape this author's eye. Every point is presented in detail, what is presented can be relied on, and uncertainties, gaps in knowledge or insufficiently exploited but promising approaches are identified and suggested as targets for further research.

Although Jennings says that he "cannot claim that [the book] is encyclopaedic", the contents cover all that can reasonably be considered relevant to fungal nutrition. Organized in 14 chapters, the text deals with fungal structures and processes (such as walls and membranes, the vacuolar compartment and translocation) as well as nutrients, which are organized into elemental groups (carbon compounds, nitrogen, metals, water and so on). The book is generally well produced and there are only a few (mainly trivial) typographical errors. The logical organization and helpful index allow specific items to be found fairly easily. Most of what might need to be known about a topic to give a vivid 'state

of the art' picture is here.

Jennings also claims that he has "avoided referring... to other eukaryotic organisms" but, thankfully, this is not entirely true. It would have been a pity if the reader had missed the opportunity to benefit from the author's expertise in and knowledge of plant physiology. Fortunately, a good deal of pertinent and interesting plant physiology appears in asides or brief comparisons.

If this self-imposed boundary is overstepped, others are carefully maintained. After mentioning contrasting ideas about fungal wall growth, for instance, Jennings baldly states that "[it] is not proposed to enter into that debate here". The reader's train of thought is thus kept firmly on the rails. There is a price to pay for this single-mindedness. The text is not an easy read. Physiology is a hard task-master, requiring a bit of mathematics here, a touch of biophysics there, a pinch of basic chemistry and, increasingly these days, a sprinkling of molecular biology. Jennings carries it all off with aplomb, but the rigorous academic demands may be off-putting to many undergraduate students, and possibly some postgraduates too. But perhaps that's as it should be. The book supports rather than supplants the teacher. The arrangement of topics will not be to everyone's taste, and teachers will probably have to devise their own favourite ways of easing the student's progress through the field. But the resource is here.

From now on, and well into the next century, lecture hand-outs, tutorial reading lists and essay study guides around the world will be liberally sprinkled with references to Jennings, 1995. At last we have a real fungal physiology text. □

David Moore is in the School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK.

Picture this

Peter Brimblecombe

Atmosphere, Climate and Change. By Thomas E. Graedel and Paul J. Crutzen. W. H. Freeman/Scientific American Library. Pp. 196. \$32.95, £19.95.

AREAS of science allied to natural history have always attracted wide public interest. Since Victorian times there have been many finely illustrated works on meteorology. But the breadth of twentieth-century atmospheric science and its emphasis on numerical modelling and chemical kinetics act as a barrier to the general reader. *Atmosphere, Climate and*

Change is a simplified version of the authors' earlier textbook (*Atmospheric Change: An Earth System Perspective*, W. H. Freeman, 1994; for a review see *Nature* 367, 695 (1994)). The new book is superbly produced, with wonderful typography, fine paper and excellent binding. It is a delight to hold. These well known scientists approach the subject with a discussion of long-term climate change and move on to atmospheric chemistry and modelling.

The writing is interesting and full of allusion to art, literature and everyday experience — all calculated to maintain the reader's interest. There are a few omissions, such as the contemporary issue of PM-10 and public health. Great care has been taken to convey scientific ideas; for example a diagram of the pH scale is marked with lemon juice, vinegar and milk. Yet this chapter contains around 30 chemical reactions embedded in the text. Although I appreciate the need for equations, I suspect that the ordinary reader will find these ones beyond comprehension. Perhaps they should have been put into separate sections that could be skipped by nonscientists.

Colour plates catch the reader's attention, yet I often found their captions — and sometimes even the pictures themselves — puzzling. Sandro Botticelli's "The Birth of Venus" does not seem the ideal illustration of the association of wind, water and sunshine and the origin and maintenance of life. Quite the contrary: this highly allegorical picture has been thought to contrast the barrenness of the sea with the civilizing fecundity of the land. I do not regard the satanic vision of Constantin Meunier's "In the Black Country" as an example of chimneys viewed with civic pride. His preoccupation with working conditions of miners and labourers makes this most unlikely. It is admirable to use art to place science within the context of human experience. But the difficulties with the illustrative material in this book should remind us that picture editing requires just as much care as the review of science.

I hope non-experts will enjoy this book. It is a wonderful introduction to the intricacies of a modern science relevant to serious questions confronting humankind. □

Peter Brimblecombe is in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

■ Cambridge University Press has just published *Basic Physical Chemistry for the Atmospheric Sciences* by Peter V. Hobbs. It aims to provide a clear and concise review of basic principles for undergraduates. Contains more than 150 exercises. £40 (hbk), £14.95 (pbk).

Structural determinants of nuclear receptor assembly on DNA direct repeats

Fraydoon Rastinejad^{*}, Thomas Perlmann^{†‡}, Ronald M. Evans[†]
& Paul B. Sigler^{*§}

^{*} Department of Molecular Biophysics and Biochemistry and The Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06510, USA

[†] Howard Hughes Medical Institute, Salk Institute, La Jolla, California 92037, USA

Nuclear receptor heterodimers recognize response elements composed of two direct repeats of the consensus sequence 5'-AGGTCA-3' separated by one to five base pairs. The 1.9 Å crystal structure of the complex formed by the DNA-binding domains of the 9-*cis* retinoic acid receptor and thyroid hormone receptor bound to a thyroid-response element shows that the subunits interact through a DNA-supported interface involving the carboxy-terminal extension of the DNA-binding domain of the thyroid hormone receptor. The stereochemistry suggests a mechanism by which heterodimers recognize the inter-half-site spacing between direct repeats.

LIPHILIC hormones such as steroids, retinoic acid, thyroid hormone and vitamin D₃ function by interacting with ligand-activated transcription factors that comprise the steroid/nuclear receptor superfamily¹. The receptors for thyroid hormone (TR), vitamin D₃ (VDR), and all-*trans* retinoic acid (RAR) form heterodimers with the 9-*cis* retinoic acid receptor (RXR) on bipartite hormone-response elements (HREs) composed of non-symmetrical head-to-tail tandem AGGTCA 'half-sites'²⁻⁷. As shown in Fig. 1a, the HREs for TR, VDR and RAR differ from one another only by the number of neutral base pairs separating the hexameric 'direct' repeats.

C-terminal to the DNA-binding domain (DBD) is a domain responsible both for ligand binding and for DNA-independent dimerization which *in vitro* allows preformation of certain dimers in solution before DNA targeting⁸. However, the DBDs of RXR, TR, RAR and VDR alone can direct the binding of hetero- or homodimers to properly spaced half-sites⁹⁻¹³. The DBDs alone do not dimerize in solution even at millimolar concentrations, but their assembly on the appropriate target DNA is highly cooperative, overwhelmingly favouring dimeric binding of the correct heterodimers to the properly spaced half-sites. This implies that the target DNA leads to the formation of a highly specific dimer interface which places the DBDs in register with the half-sites of their respective response elements. For example, the DBDs of RXR and TR bind with high cooperativity and affinity to a TREdr4, but will bind weakly and without cooperativity to similar half-sites spaced either three (VDREdr3) or five (RAREdr5) base pairs apart (Fig. 1a). This DNA-mediated dimerization among the nuclear receptor DBDs is reminiscent of that established for glucocorticoid and oestrogen receptors (GR and ER) by biochemical and X-ray crystallographic studies¹⁴⁻¹⁶. However, the asymmetrical head-to-tail arrangement implied by the tandem of direct repeats calls for a different dimerization mechanism from that seen in the symmetrical binding of GR and ER on the inverted repeats of their respective response elements.

To understand the stereochemical principles underlying discrimination of HREs by the size of the inter-half-site spacer and the assembly of the nuclear-receptor dimers, we have solved the crystal structure of the heterodimer formed by RXR and TR

DBDs on a TRE composed of a tandem of half-sites separated by four base pairs (TREdr4). This structure, refined at 1.9 Å, shows the critical features of protein-protein and protein-DNA interactions that specify the co-targeting of retinoid and thyroid receptors to the TREdr4. Because the three-dimensional structures of the DNA-bound DBDs from the GR, ER, TR and RXR are remarkably conserved (Fig. 2c), we have modelled the interfaces, yet to be observed directly, that would be expected to form on repeats with spacing other than four base pairs. To do this, we changed the number of base pairs in the inter-half-site spacer while maintaining the specific interface to the DNA half-sites, and used molecular dynamics and energy minimization to produce the dimer interfaces likely to be formed between the DBDs of other nuclear receptors. These models are given credence by the fact that with no specific intent, unique and non-conserved residues form the dimer interface. Here we describe both the crystallographically observed RXR-TR DBD interactions on their target DNA as well as the dimerization interfaces likely to be produced between other receptor dimers, derived from modelling.

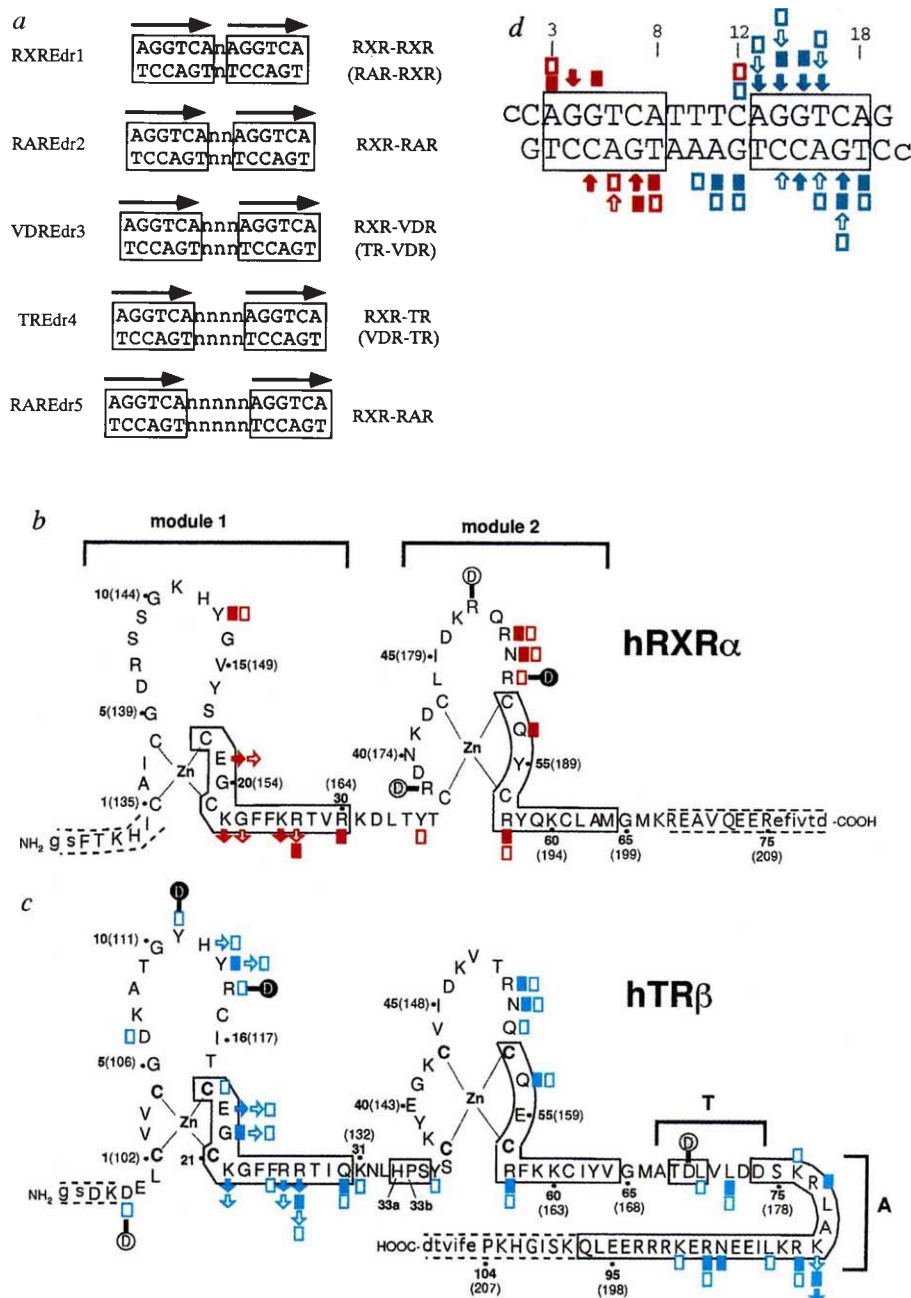
Structure determination and refinement

An 80-residue fragment of the human RXR- α (residues 130-209, Fig. 1b) and a 111-residue fragment of human TR- β (residues 99 to 207; Fig. 1c), which together bind with high cooperativity and selective affinity to the TREdr4⁹, were co-crystallized with the TREdr4 shown in Fig. 1d. Orthorhombic crystals (space group $P2_12_12_1$, $a=40.9$, $b=65.6$, $c=125.7$) diffracted to 2.7 Å on laboratory sources and to beyond 1.9 Å under a stream of cold nitrogen (-167 °C) using synchrotron radiation ($\lambda=0.91$ Å). The asymmetric unit of the crystal consists of one DNA duplex and one molecule each of the TR DBD and the RXR DBD. Phases were derived by multiple isomorphous replacement (MIR) using derivatives in which one or two thymines were replaced by 5-iodouracil. An initial 2.7 Å MIR electron-density map was readily interpretable in terms of the DNA duplex, backbone and side chains of TR for residues 99 to 200 (-3 to 97 in the common numbering scheme), and all backbone atoms and some side chains of RXR for residues 135 to 200 (1 to 67 in the common numbering scheme, which will be used here). This partial model was refined by least-square minimization and slow cooling (X-PLOR)¹⁷, initially with tight restraints for tetrahedral zinc coordination and Watson-Crick DNA base pairing, resulting in an *R*-factor of 29% for all data to 3.0 Å. Calculated and experimental phases were combined (Sigma-A, part of CCP4

[‡] Present address: Department of Cellular and Molecular Biology, Medical Nobel Institute, Karolinska Institute, S17177 Stockholm, Sweden.

[§] To whom correspondence should be addressed.

FIG. 1 a, The 'spacer rule' for nuclear-receptor assembly on DNA direct repeats (dr). RXREdr1, RAREdr2, VDREdr3, TREdr4 and RAREdr5 refer to 9-*cis* retinoic acid, vitamin D₃, thyroid hormone, and all-*trans* retinoic acid receptor DNA-response elements, each composed of direct repeats with 1, 2, 3, 4 or 5 base pairs separation, respectively. RXR occupies the upstream half-site in these complexes^{9,10,36}; its dimerization partner is indicated on the right. Other heterodimeric arrangements are seen on direct repeats of the same half-sites^{37,38}, and these heterodimers and their polarities are also indicated in parentheses on the right. b–d, Sequences of molecules used for crystallographic analyses. b, Human RXR- α DBD; and c, the human TR- β showing the Cys–Zn coordination and α -helices, which are boxed. The numbering in bold (and from hereafter) is relative to the first Zn-coordinating cysteine of the DBD, whereas the numbers in parenthesis relate the residue position relative to the N terminus of the full-length receptors^{39,40}. The DBDs are composed of two modules, which are indicated by brackets. Solid arrows indicate residues that make direct base contacts in the RXR (red) and TR (blue), respectively. Solid rectangles indicate residues that make direct phosphate contacts for RXR (red) and TR (blue), respectively. Open symbols indicate water-mediated hydrogen bond to the bases (open arrows) and the phosphates (open rectangles). We define here two structural elements of TR, the A-helix comprising residues 74–97, and the T region, which is the connector loop (residues 67–74) between the A-helix and the TR core DBD. Note that a T box has been previously defined for RXR (residues 67–78), and an A-box for NGFI-B²⁰, which partly correspond to the designations used here for TR. Residues that make the dimer interface are shown as ovals, with the shading indicating the pairings between residues that form the interaction between RXR and TR DBDs. Disordered regions at the N and C termini are indicated by the dashed lines. Cloning artefacts from the expression vector for the RXR and TR DBDs are indicated by small letters. A two-residue kink is seen in the middle of the A α -helix and is caused by crystal-packing contacts with the C terminus of a symmetry-related α -helix. d, The sequence of the DNA used in crystallization. The consensus half-sites are boxed. The spacer sequence is derived from the TREdr4 found near the mouse leukaemia virus (MLV) promoter⁹. The



software) to produce an improved map from which additional RXR DBD residues were added. Tight restraints on the Zn–S coordination and base pairing were released, and multiple rounds of rebuilding and refinement with X-PLOR, including refinement of individual B-factors, led to a final R-factor of 20.6% for all data between 6.0 and 1.9 Å (free R-factor 26.9%) with good stereochemistry. The final model contains residues 99 to 201 of TR and residues 135 to 201 of RXR, 18 base pairs, and 230 water molecules. The quality of the data, phases and refinement are summarized in Table 1 and a region of the electron-density map ($2F_o - F_c$) is shown in Fig. 3b.

Overall architecture

The TREdr4 DNA accommodates the RXR–TR DBD heterodimer as B-form DNA without significant distortion. The four

symbols indicate contacts with RXR (red) and TR (blue) as in a and b. The cytosines at the 5' end of each strand, shown in small letters, are not seen in the structure.

intervening base pairs in the DNA response element place the centres of the hexameric repeats ten base pairs apart, nearly one full turn of the DNA helix. The bound DBDs engage the major grooves of successive repeats with the same geometry, placing them on the same face of the TREdr4 target. The disposition of the subunits with respect to the half-sites resembles that seen in the 'specific' half of the GR–DBD complex with DNA¹⁵ and both halves of the ER–DBD–DNA complex¹⁶. However, the tandem arrangement of the half-sites in TREdr4 imposes a head-to-tail orientation of the DBDs of RXR and TR, which has no rotational symmetry (Fig. 2a, b). Although the two half-site consensus sequences of TREdr4 are identical, the RXR DBD occupies the upstream DNA element, and the TR DBD the downstream DNA half-site, in agreement with biochemical

data^{9,11,12,18}. This polarity is a consequence of a unique stable heterodimer interface formed only in the observed direction.

Protein tertiary structures

Common elements of nuclear receptor DNA-binding domains.

The first 66 residues or 'core' of the DBD in the steroid/nuclear receptor family are highly conserved. The DBDs of TR and RXR in this region are 52% identical in sequence, and have a remarkably similar overall fold (Fig. 2c). TR contains a two-residue insertion (His 33a-Pro 33b) which is not present in RXR, GR or ER. But these residues, together with Ser 34, form a 3_{10} -helix turn which compensates for the increased contour length and maintains a constant distance between the recognition helix and Zn module 2, thus preserving the overall fold seen in RXR, GR and ER (Figs 1c, 2c). The N-terminal modules of both RXR and TR coordinate their zinc atom in the *S* configuration, whereas the C-terminal modules are coordinated in the *R* configuration.

There are interesting differences between the X-ray structure of the RXR DBD bound to DNA and the NMR structure of the uncomplexed RXR DBD residues 1–89 (ref. 19). First, the NMR structure of the RXR DBD shows an inverted chirality (the *S* configuration) in the second zinc module. Second, the sequence beyond the conserved Gly-Met at positions 66–67 of RXR, reported to be an α -helix by NMR¹⁹, is not visible in our structure because of crystallographic disorder. This region (residues 67–78) is called the T box and plays a role in cooperative binding of RXR homodimers to direct repeats with a one-base-pair spacer^{10,20}.

Structural elements in the C-terminal segment of the TR DBD.

The sequence of TR C-terminal to the core DBD (beyond residue Met 66) forms structural elements not seen previously in nuclear receptors^{15,16,19,21}. TR residues 66–73, corresponding to the T box of RXR, form a connecting loop interrupted by a 3_{10} -type helical turn that lies perpendicular to the long axis of DNA.

Asp 69 from this region of TR makes a key contact with Arg 38 of RXR to establish part of the inter-subunit interface (Fig. 4a). The TR sequence starting at residue 74 then forms a long α -helix (A-helix) composed of 24, mostly charged, residues (Fig. 1c). This helix contains the region (residues 79–85) that in nerve

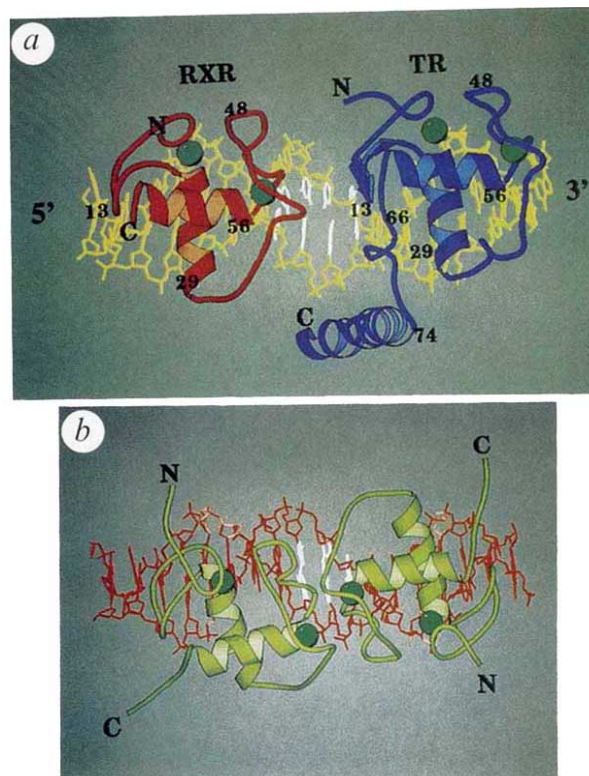
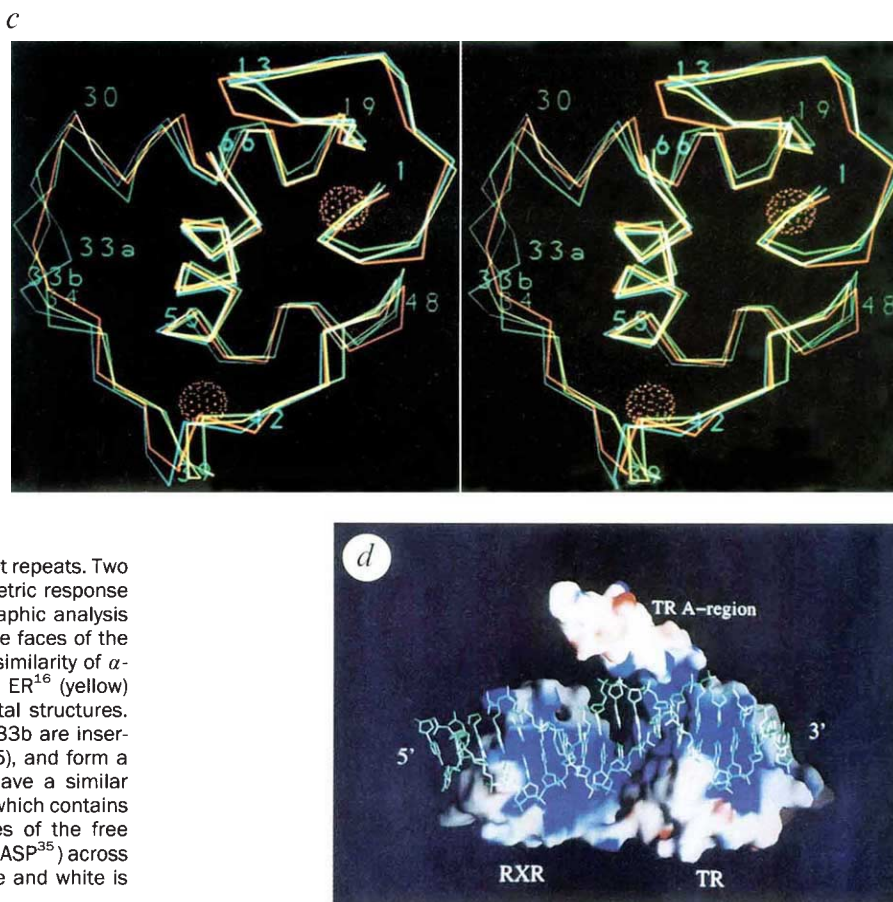
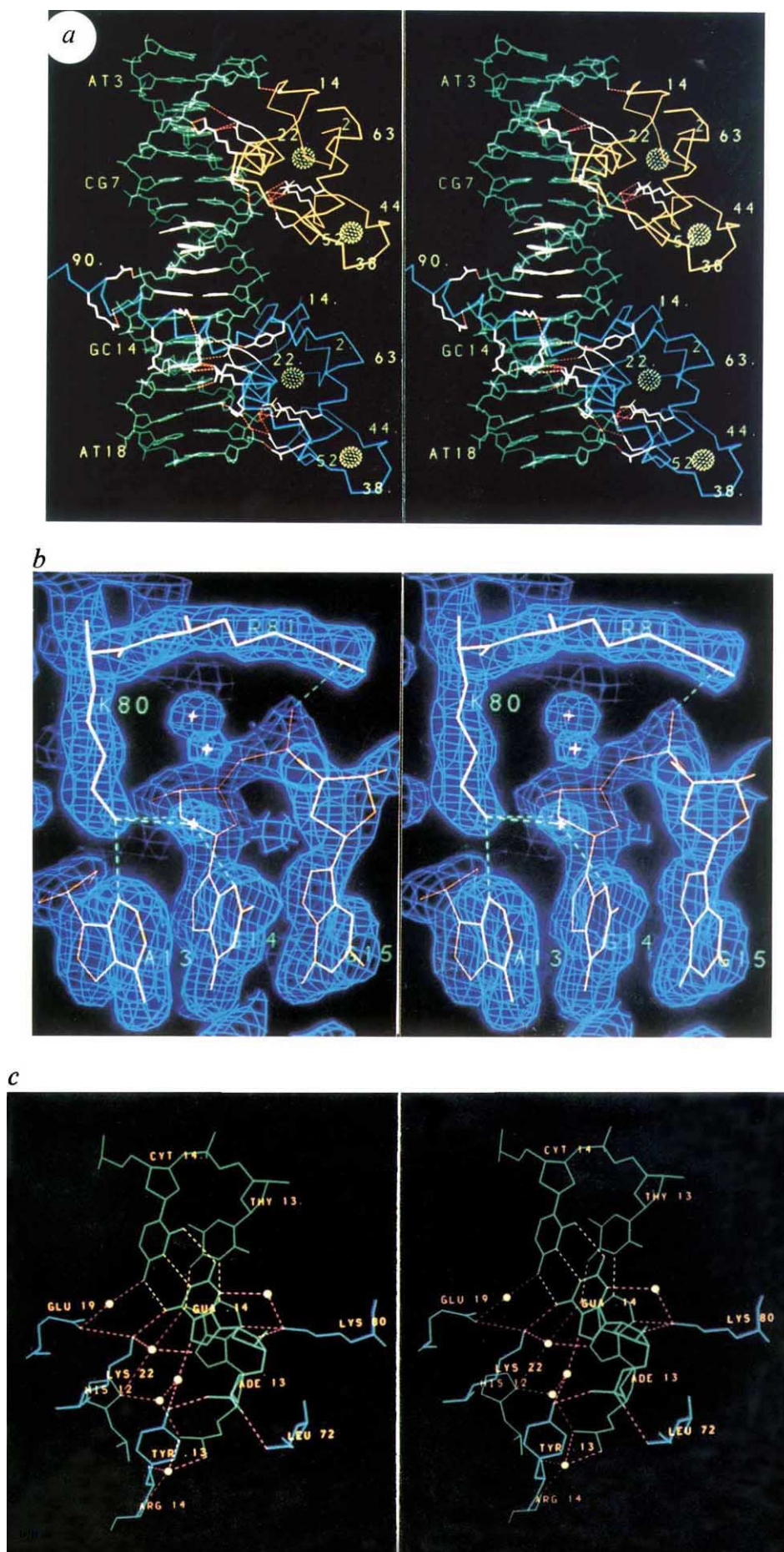


FIG. 2 a, The overall architecture of RXR (red) and TR (blue) DBDs on a direct repeat sequence of AGGTCA with 4-bp separation, TREdr4 (yellow). The 5' and 3' ends of the DNA duplex are indicated. b, GR-DBD symmetrical homodimer (green) on a palindrome repeat of AGAACA with 3-bp separation. Spacer base pairs are white; Zn atoms are green spheres. Number is relative to the first coordinating cysteine of the DBDs. Because of the polarity of the RXR-TR interactions, residues from the upstream subunit (RXR) that contribute to the dimer interface are distinctly different from those that would be presented by the downstream subunit (TR). Consequently, the observed polarity cannot be reversed without loss of favourable intersubunit contacts. Strong symmetric contacts, such as those of the GR homodimer or ER homodimer, are not possible at intersubunit interfaces formed on direct repeats. Two TR DBD molecules could bind independently on a symmetric response element without a spacing and, as expected, crystallographic analysis shows the subunits in this arrangement to lie on opposite faces of the DNA helix without any intersubunit contacts⁴¹. c, overall similarity of α -carbon traces of RXR (red), TR (blue), GR¹⁵ (green) and ER¹⁶ (yellow) core DBDs (first 66 residues) derived from DNA co-crystal structures. The Zn positions are red spheres. TR residues 33a and 33b are insertions not common to other nuclear receptor DBDs (Fig. 5), and form a 3_{10} helix with residue 34. Notice that all four DBDs have a similar structure in the second Zn module from residues 42–55, which contains a disordered helix not well represented in NMR studies of the free DBDs²¹. d, Electrostatic surface potential (displayed by GRASP³⁵) across the RXR-TR heterodimer; blue is positive, red is negative and white is neutral.





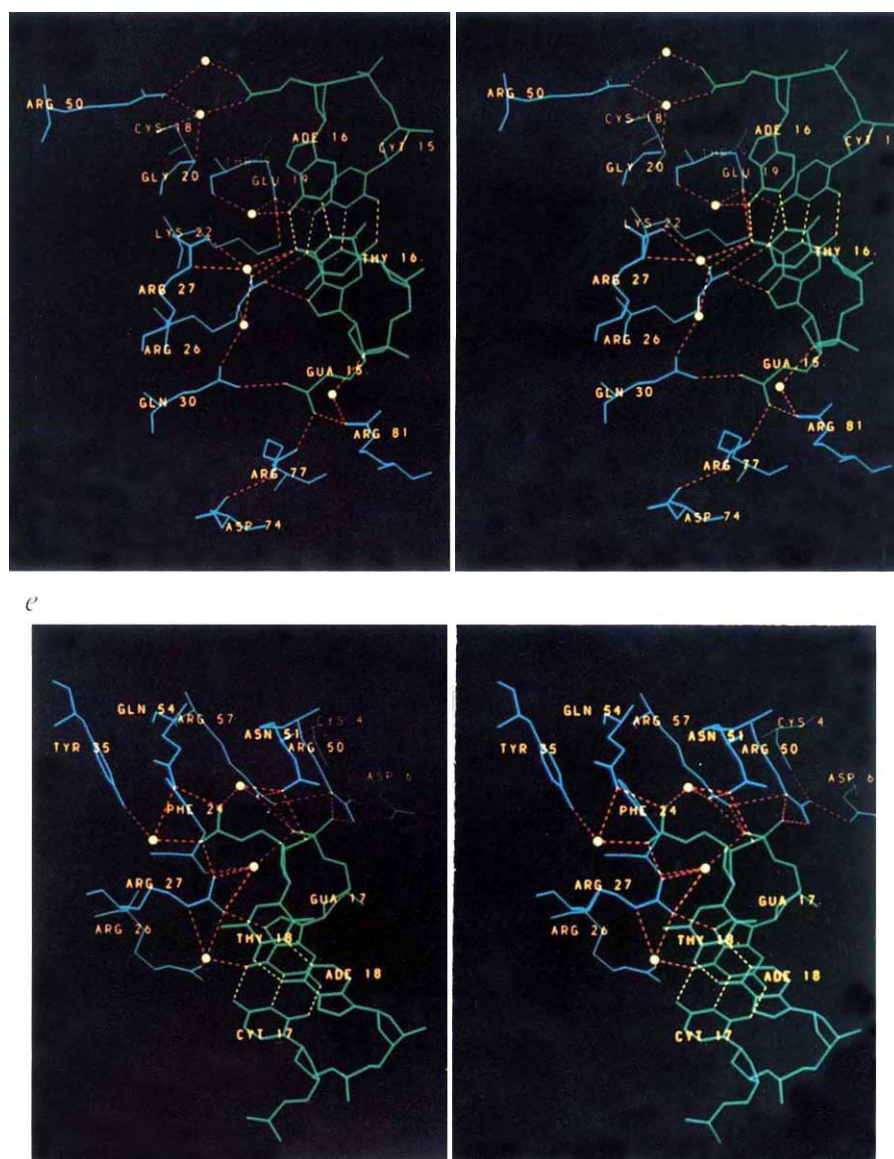


FIG. 3 *a*, Stereo view of the backbone traces of the DBDs of RXR (orange) and TR (blue) on TREdr4. Only side chains involved in direct base or phosphate contacts are shown. The dotted red lines indicate hydrogen bonds. The residues are relative to the first Zn-coordinating cysteine; Zn atoms are yellow and spacer DNA is white. Residues 92–97 of TR DBD are not shown. *b*, Stereo view of the electron-density map ($2F_o - F_c$) showing the interactions of TR Lys 80 and Arg 81 with the backbone and minor-groove functional groups of DNA positions

A(13), G(14) and G(15); (+) indicates water molecules. *c–e*, Details of the specific interactions in the TR DBD interface with its half-site DNA, showing both direct and water-mediated contacts along the six base pairs. The view is along the DNA axis towards the 5' end. Water molecules are yellow spheres, amino acids are blue, and base pairs are green. Hydrogen bonds between base pairs are indicated by dotted yellow lines, and those between the TR DBD and DNA are shown in dotted red lines.

growth factor inducible-B (NGFI-B) is called the A-box because it binds adenines 5' to the single DNA half-site of NGFI-B²⁰. The TR C-terminal α -helix makes no tertiary contacts with the rest of the DBD but instead projects across the minor groove, where it makes an extensive interface with the DNA. As shown in Fig. 3*a* and *b*, its orientation is fixed by contacts with the phosphates of both strands and by interactions between Lys 80 and the minor-groove N3 atoms of bases A13 and G14, the first two base pairs of the downstream half-site (Fig. 3*b*, *c*). The regions corresponding to this α -helix in VDR and GR contain the nuclear localization signals^{22,23} and thus nuclear localization may be an additional function of this structure in TR.

The protein–DNA interface

The complete protein–DNA interface is extensive, burying 3,357 Å² of water-accessible surface, two-thirds of which is

buried by the TR DBD/DNA half-site interaction. The A-helix of TR significantly increases the TR interface with DNA through direct and water-mediated hydrogen bonds with the phosphates and the minor groove edges of base pairs (Fig. 1*c*, *d*). The entire face of the dimer that interacts with DNA contains a positive electrostatic charge potential that complements the negative electrostatic potential of the DNA (Fig. 2*d*). The recognition α -helices within each of the DBDs (residues 19–30) engage the major-groove side of the DNA half-sites, making direct contacts with the base pairs and backbone-phosphate oxygens (Figs 1 and 3*a–e*) that resemble those seen in the ER DBD–DNA structure¹⁶. The extensive interactions between the A-helix of the TR DBD with the minor groove of the half-site and spacer (summarized in Figs 1*c* and 3) appear to stabilize the position of the TR core DBD so as to make a more extensive interface with the major groove of the downstream half-site of the

TREdr4. As a result, the DNA contacts made by the recognition helix of the TR DBD (residues 18–30) far exceed those made by the DBDs of RXR, ER and GR to their respective half-sites.

In view of the more extensive interactions of the TR DBD with the TRE half-site and upstream spacer nucleotides (Fig. 1d), it is not surprising that, unlike other nuclear-receptor DBDs, the TR DBD binds DNA efficiently as a monomer, and recognizes DNA sequences at the 5' side of its consensus half-site^{9,24,25}. The phosphate and minor-groove contacts formed by the A-helix of TR DBD (Fig. 1d) account for the interference with N3 methylation at TREdr4 positions 11, 12 and 13 (Fig. 1d)¹⁸. These interactions extend the contact surface of the TR DBD DNA to beyond the consensus 6-base-pair half-site. Inter-

estingly, some orphan receptors, such as NGFI-B and embryonal long terminal repeat-binding protein (ELP) are able to bind efficiently as monomers. Mutagenic analysis experiments indicate that such receptors also employ a sequence segment C-terminal to the core to interact with bases (in this case specific sequences), that, like those of the TRE, are contiguous with the 5' end of a TRE-like half-site^{20,26}.

Residues mediating subunit contacts also contribute to the stability of the protein–DNA interface, as most of the residues mediating dimerization contacts are themselves involved in DNA contacts, and all are clustered with neighbouring residues that make phosphate contacts (Figs 1b, c, 3c–e). This linkage of DNA contacts with intersubunit contacts produces cooperative

TABLE 1 Summary of crystallographic analysis

Phase determination										
Resolution shell (Å)	8.39	6.45	5.24	4.41	3.81	3.35	2.99	2.70	Total	
Mean figure of merit*	0.78	0.87	0.82	0.75	0.58	0.52	0.36	0.31	0.509	
I-9,11†										
Number of unique reflections	201	354	585	804	1,041	1,346	472	0	4,803	
Phasing power‡	1.29	2.31	1.99	1.57	0.88	0.85	0.79	—	1.20	
I-11†										
Number of unique reflections	218	390	608	856	1,105	1,424	1,755	2,185	8,541	
Phasing power‡	0.86	1.42	1.44	0.98	0.71	0.42	0.35	0.56	0.69	
I-18†										
Number of unique reflections	211	373	588	819	1,053	1,357	15	0	4,416	
Phasing power‡	0.80	1.21	1.07	0.86	0.66	0.67	0.62	—	0.79	
I-3,13†										
Number of unique reflections	221	394	612	855	1,102	1,418	1,762	1,916	8,280	
Phasing power‡	1.05	1.65	1.51	1.10	0.64	0.55	0.64	0.58	0.84	
I-13†										
Number of unique reflections	221	393	610	858	1,100	1,388	660	0	5,230	
Phasing power‡	0.90	1.74	1.59	1.16	0.59	0.57	0.44	—	0.85	
Parent data and refinement§										
Resolution shell (Å)	3.45	3.02	2.74	2.54	2.39	2.27	2.09	1.96	1.90	Total
Number of reflections used in refinement (6–1.9 Å)	8,421	4,211	4,128	4,102	4,116	4,015	7,767	5,940	2,175	44,875
Per cent coverage	98	98	97	96	95	94	90	69	51	87
I/σ	21.9	21.4	18.6	17.6	15.1	12.4	9.6	5.5	4.1	18.9
R _{sym}	0.074	0.083	0.100	0.113	0.124	0.149	0.169	0.234	0.311	0.083
R-factor (all data)										20.6
Free R-factor (all data) ³⁰	0.231	0.249	0.267	0.284	0.303	0.286	0.303	0.341	0.391	26.9
R.m.s. deviations										
										bond length 0.020 Å
										bond angle 1.969°

RXR and TR DBDs were overproduced, using *E. coli* clones, as glutathione S-transferase (GST) fusion proteins, and purified using glutathione–Sephadex affinity chromatography. GST was subsequently removed by thrombin proteolysis followed by Mono-S FPLC. Oligonucleotides were synthesized by the phosphoramidite method and purified with the 5' dimethoxytrityl blocking group attached on a reverse-phase C4 HPLC column. After acid detritylation, oligonucleotides were repurified in the same way. The two DNA strands were annealed by raising the temperature to 90 °C and cooling to room temperature over 3 h. The protein–DNA complex was made using stoichiometric equivalents of RXR DBD, TR DBD, and duplex DNA, and concentrated to 0.6 mM (each species) in 25 mM imidazole buffer, pH 7.0, 400 mM NH₄Cl, 1 mM DTT, 5 mM MgCl₂. Crystals were grown using hanging drops at 17 °C by the addition of 5 µl of this complex to an equal volume of the precipitant solution (40% w/v Peg-3350, 25 mM imidazole buffer, pH 7.0, 400 mM NH₄Cl, 10 mM DTT, and 5 mM MgCl₂). Diffraction quality crystals (0.1 × 0.3 × 0.5 mm) in the space group P2₁2₁2₁ grew within one week. Crystals were streaked through a cryosolvent containing the precipitating solution adjusted to contain 30% (v/v) glycerol and then flash-frozen in liquid propane. The parent data were collected on a CCD detector at the A1 beamline of the Cornell High Energy Synchrotron Source (λ = 0.91 Å; CHESS); the remaining data sets were collected on a RAXIS II using copper Kα from a RU200 generator (mirror focused). All data were processed with DENZO (Z. Otwinowski) and scaled with SCALEPACK (Z. Otwinowski). The parent and derivative data sets were placed on the same relative scale using SCALEPACK. Heavy-atom derivatives were obtained by growing crystals containing DNA with 5-iododeoxyuridine (5-IdU) substitutions for thymidines. Iodine positions, found using difference Pattersons and confirmed with cross-difference Fourier, were refined using ML-PHARE³¹. The polypeptides and the DNA were traced using the programs FRODO³² and O³³. Refinement, using X-PLOR¹⁷, consisted of cycles of rigid-body, least-squares and simulated-annealing steps. Throughout refinement, the model was readjusted with the aid of difference maps (including 'omit' maps), causing the free R-factor to fall continuously. Water sites were added in the later stages, using (F_o – F_c) and (2F_o – F_c) maps. All water molecules have a B value of <70 Å² and participate in at least one hydrogen-bonded interaction. Graphics presented here used the programs FRODO³², Molscript³⁴ and Grasp³⁵.

* Mean figure of merit is $\int P(\phi) \exp(i\phi) d\phi / \int P(\phi) \exp(i\phi)$, where P is the probability distribution of the phase ϕ angle.

† The derivatives contain 5-IdU substitution at the positions indicated; for example, I-9,11 refers to substitutions at positions 9 and 11 (Fig. 1c).

‡ Phasing power is calculated as $\sum |F_H| / \sum [|F_H| \exp(i\phi_c) + F_H^* - |F_H^*|]$, where F_H^* is the heavy-atom structure factor, $|F_H|$ and $|F_H^*|$ are observed amplitudes for the parent and heavy-atom derivative, respectively, and ϕ_c is the calculated phase; the summation is over all observations.

§ The data set used in refinement contains a single 5-IdU substitution at position 11 (Fig. 1c). Friedel mates were kept separate, and refinement was carried out against both (where measured) using the real and imaginary scattering factors for zinc and iodine.

|| $R_{\text{sym}} = \sum_n \langle |I_h - I_n| \rangle / \sum_n I_h$, where $|I_h - I_n|$ is the absolute difference between a reflection I_h and I_n , the average of its symmetry and Friedel equivalents.

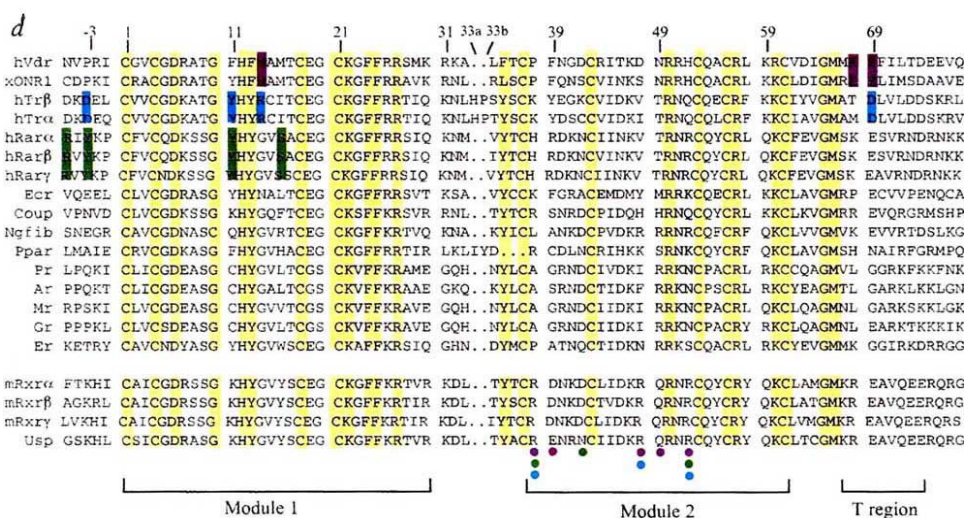
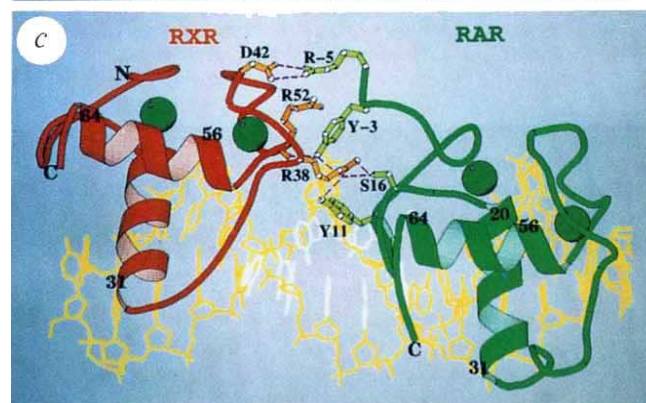
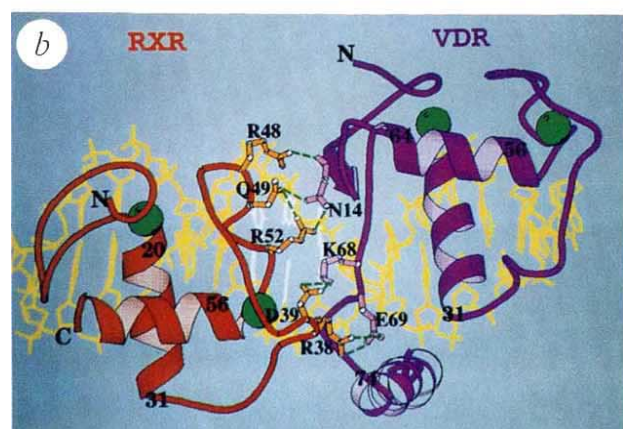
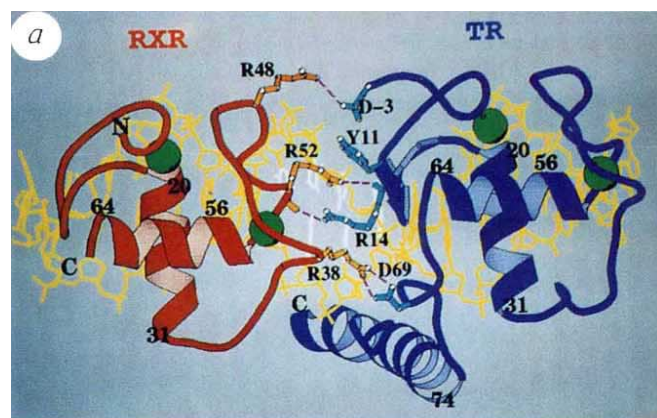


FIG. 4 Subunit interfaces on TREdr4, VDREdr3 and RAREdr5 targets respectively. Amino-acid residue numbers are relative to the first Zn-coordinating cysteine. White base pairs are the spacers; green spheres are zinc ions. **a**, The dimer interface of the RXR DBD (red) and the TR DBD (blue) seen in the crystal structure. Only residues participating (**a**) or expected to participate from computer modelling (**b**, **c**) are shown. Hydrogen bonds between subunits are shown as dotted lines. In the models shown in **b**, **c**, the protein structure is allowed to readjust to minimize the overall energies calculated from van der Waals and electrostatic interactions (including those with DNA) using positional and simulated annealing refinement (X-PLOR)¹⁷. **b**, The proposed interface between RXR (red) and VDR (purple) on VDREdr3; and **c**, the proposed interface between RXR and RAR (green) on RAREdr5. **d**, Amino-acid sequences of various nuclear-receptor DBDs⁴² (and their isotypes for TR^{39,43}, VDR^{44,45}, RAR⁴⁶⁻⁴⁹ and RR^{40,50}) showing those that participate in the dimer interface. Dimerization contacts described in Fig. 3a-c are indicated here by colour shading for TR (blue), VDR (purple) and RAR (green). Yellow indicates sequence identities; circles represent RXR residues that form the corresponding dimerization contacts with TR (blue), VDR (purple) and RAR (green). m, Mouse; h, human; x, *Xenopus*. The three RXR isoforms α , β and γ , together with the homologous protein from *Drosophila* (Usp) are shown in the lower part of the figure. xONR1 is a receptor related to VDR.

binding of the TR and RXR DBDs on the TREdr4 (see discussion below).

Of potential interest to the behaviour of the DBDs on direct repeats is the network of hydrogen bonds formed by TR Arg 26, which is found in most nuclear but not steroid receptors. Figure

3*d, e* shows that TR Arg 26 donates direct hydrogen bonds to DNA bases at positions G15 and T16 and water-mediated hydrogen bonds to DNA positions T16 and G17. Substitution of an alanine for Arg 26 causes a 20-fold reduction in the affinity of the TR DBD (alone) for naturally occurring TREs²⁷.

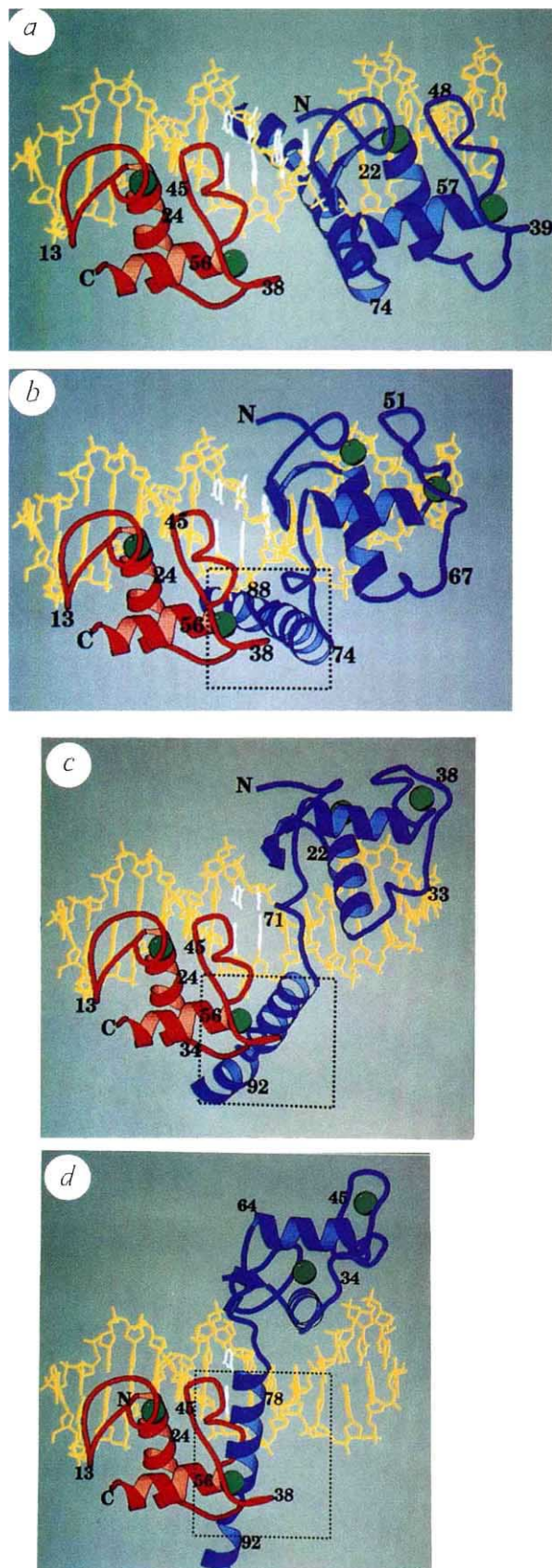
The dimer interface

Contacts between the RXR DBD and the TR DBD straddle the minor groove of the spacer. Unlike the symmetric case of ER or GR, the head-to-tail subunit arrangement brings different regions of each protein to the dimerization interface. Residues from the second zinc module of the RXR DBD are positioned at the upstream side of the interface, and interact with residues from the first zinc module and the T region of the TR DBD (Fig. 4*a*). The interactions from the RXR DBD are formed by three arginines: Arg 38, Arg 48 and Arg 52, the first two of which make salt bridges with TR Asp 69 and Asp (-3), respectively (Fig. 4*a*). Asp -3 is located in a region preceding the core DBD, which was predicted to be in the heterodimerization interface from biochemical studies¹¹. RXR Arg 48 and Arg 52 lie within a distorted α -helical region (47–53; Figs 1*b* and 2*a*) which in the GR and ER participate in symmetrical dimer interfaces. This distorted helical secondary structure is, however, either not visible or is poorly distinguished in NMR structures of the non-complexed DBDs^{19,21} (Fig. 2*c*).

Figure 1*b, c* shows that the residues mediating the dimerization contacts are clustered in regions that make phosphate contacts. The interactions of Arg 52 of RXR highlight the interdependence of DNA binding and dimer assembly. Arg 52 makes a pair of reciprocal interactions with Arg 14 of TR, where a terminal guanidino NH from each side-chain donates a hydrogen bond to its partner's backbone carbonyl oxygen. In addition, RXR Arg 52 makes a planar stacking interaction²⁸ with the Tyr 11 of TR, further reinforcing the dimer interface. The guanidino group of RXR Arg 52 is buttressed by a water-mediated hydrogen bond with a backbone oxygen of the DNA at position C12 in the spacer sequence, immediately preceding the TR half-site. RXR residues surrounding Arg 52, notably Arg 50, Asn 51 and Gln 54, form hydrogen bonds to the DNA backbone and thereby support its proper position in the RXR dimer interface (Fig. 1*b*). Thus the DNA contacts support the dimer interface, which in turn supports both the specific and nonspecific contacts with the DNA.

These mutually supporting interactions between subunits and DNA backbone reinforce a lesson learned from the steroid receptor–DNA complexes concerning the importance of inter-half-site spacing²⁹; namely, the DBDs of the nuclear receptors, which are monomers in the absence of DNA and which generally bind weakly as monomers to a single half-site, bind with high affinity and strong cooperativity to response elements in which the appropriate half-sites are spaced correctly. Cooperative enhancement of protein–protein and protein–DNA interactions at the dimer interface places the subunits in registration with the specific contact surface of the half-site and reinforces the 'specific' recognition.

It is important to note that residues contributing to the dimer interface occupy non-conserved positions (Fig. 4*d*), and are relatively specific to a particular DBD and its isotypes, thereby



◀ FIG. 5 The steric role of the TR C-terminal α -helix (residues 74 to the C terminus) in spacer discrimination. The A-helix retains all of the structure and DNA contacts that fix its orientation on DNA relative to the core, as observed crystallographically on TREdr4. The RXR (red) and TR (blue) DBDs as observed in the crystal structure on a, TREdr4, and modelled on b, VDRdr3; c, dr2; and d, dr1. Dotted boxes show regions of steric clash involving the TR A-helix. On dr1 and dr2, main-chain atoms on the TR A-helix and the second Zn module of RXR would clash, restricting co-occupancy by RXR and TR. On VDREdr3, side-chain atoms of the TR A-helix would clash with the second RXR Zn module.

ensuring a unique mode of dimerization on the corresponding response element. This point will be addressed in connection with modelling of the heterodimeric interfaces formed on VDREdr3 and RAREdr5.

Spacer discrimination by receptor DBDs

RXR forms heterodimers with VDR, TR and RAR on DNA targets that differ only by their inter-half-site spacing (Fig. 1a). To model the heterodimeric complexes with VDR and RAR spaced three and five base pairs apart, respectively, we have readjusted the RXR-TR DBD interface seen in the crystal structure by changing the number of intervening base pairs by one. The validity of this approach rests, in part, on the nearly identical conformation of the core DBDs of RXR, TR, GR and ER, and their nearly identical alignment relative to their respective half-sites. The models were made by fixing the orientation of the RXR DBD and its associated half-site, and rotating and translating the complex formed by the TR DBD and its half-site as a unit, so as either to reduce (VDREdr3) or to increase (RAREdr5) the inter-half-site spacing by one base pair. All TR residues were then replaced by those of the VDR or RAR. Energy minimization was carried out as described in Fig. 4 legend. It should be stressed that no effort was made to bias the computation so that a particular substructure or residue would be included at the interface.

Figure 4b shows that RXR and VDR modelled on a VDREdr3 can make a number of favourable contacts at their interface. The key residue mediating heterodimerization for VDR is Asn 14, which makes side-chain contacts to RXR Gln 49 and Arg 52, and a backbone carbonyl hydrogen bond with RXR Arg 48. Moreover, Asn 14 is unique to VDR, setting it apart from the other receptors as the one that selects VDREdr3 as its co-target with RXR. This position is typically occupied by a glycine in other nuclear receptors, except in TR, where Arg 14 is also involved in subunit dimerization (Fig. 4d). Mutagenesis supports the role of Asn 14 in spacer discrimination by VDR¹³. The model shows that VDR Lys 68 and Glu 69 support the dimer interface by making salt bridges with RXR residues. As shown in Fig. 4b, the T region modelled for VDR has a different backbone conformation from that of TR. This change satisfies the hydrogen bonding of VDR Lys 68 and Glu 69 with RXR, and also readjusts the presumed A-helix so that it does not sterically interfere with dimerization (Fig. 5b). This readjustment also places the VDR A-helix in a registration favourable for hydrogen bonding between the basic residues and the DNA backbone.

The modelled contacts between RXR and RAR on RAREdr5 are shown in Fig. 4c. A central interaction is the hydrogen bonding between RAR Ser 16 and RXR Arg 38. Ser 16 is unique to the isotypes of RAR. This position is highly variable in nuclear receptors (Fig. 4d), but typically occupied by a hydrophobic residue which would cause a clash with RXR Arg 38. RXR Arg 38 is particularly well buttressed in this interaction, as it is adjacent to a pair of RAR tyrosines, Tyr 11 and Tyr (-3), which can hydrogen-bond through their side-chain oxygens or make planar stacking interactions with this arginine. In addition, RAR Arg (-5) can make a pair of hydrogen bonds with RXR Asp 42. These results are consistent with the fact that substitution of the pre-finger region in RAR by the corresponding segment in ER supports cooperative heterodimerization on RAREdr5 (ref. 11). The RAR Arg (-5), when replaced by a lysine from ER can also form a hydrogen-bonded ion pair with Asp 42 of RXR; and Tyr (-3), which when replaced with a threonine from ER can also accept hydrogen bonds from the side chain of Arg 38 in RXR.

RXR and TR jointly bind to TREdr4 with high specificity and cooperativity partly because they are able to make the contacts shown in Fig. 4a while reinforcing their DNA contacts. In addition, as shown in Fig. 5, the long A-helix of TR also plays a key role in spacer discrimination, as it produces sterically unfavourable interfaces on half-site spacings with less than four base pairs. Spacer discrimination on the basis of steric hindrance was suggested by *in vitro* affinity studies with chimaeric DBDs containing the C-terminal extension of TR-DBD^{11,18}. This misalignment of the A-helix and the steric clashes shown in Fig. 5b-d would suggest that the corresponding A-regions of RAR, VDR and RXR assume a different conformation or orientation that does not cause steric interference with their partner subunits, while allowing favourable contacts with the DNA backbone.

This work provides firm evidence that a DNA-supported asymmetric dimerization interface located within the DBDs of RXR and TR, and by extension, RAR and VDR, provides the molecular basis for receptor heterodimers to distinguish between closely related response elements. The structural analysis also reveals how a large repertoire of binding surfaces is generated in a protein superfamily from comparatively few amino-acid substitutions. Ultimately, a thorough understanding of the use of 'spacer' DNA as an identity element will require crystallographic analysis of a larger sampling of tandem interactions of the nuclear receptors with their cognate response elements. □

Received 3 January; accepted 15 March 1995.

- Evans, R. M. *Science* **240**, 889-895 (1988).
- Leid, M., Kastner, P. & Chambon, P. *Trends biochem. Sci.* **17**, 427-433 (1992).
- Kliwer, S. A., Umesono, K., Mangelsdorf, D. J. & Evans, R. M. *Nature* **355**, 446-449 (1992).
- Marks, M. S. *et al.* *EMBO J.* **11**, 1419-1435 (1992).
- Bugge, T. H., Pohl, J., Lonnoy, O. & Stunnenberg, H. G. *EMBO J.* **11**, 1409-1418 (1992).
- Yu, V. C. *et al.* *Cell* **67**, 1251-1266 (1991).
- Umesono, K., Murakami, K. K., Thompson, C. C. & Evans, R. M. *Cell* **65**, 1255-1266 (1991).
- Rosen, E. D., Beninghof, E. G. & Koenig, R. J. *J. biol. Chem.* **268**, 11534-11541 (1993).
- Perlmann, T., Rangarajan, P. N., Umesono, K. & Evans, R. M. *Genes Dev.* **7**, 1411-1422 (1993).
- Zechel, C. *et al.* *EMBO J.* **13**, 1425-1433 (1994).
- Zechel, C., Shen, X. Q., Chambon, P. & Gronemeyer, H. *EMBO J.* **13**, 1414-1424 (1994).
- Mader, S. *et al.* *EMBO J.* **12**, 5029-5041 (1993).
- Towers, T. L., Luisi, B. F., Aslanov, A. & Freedman, L. P. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6310-6314 (1993).
- Hard, T. *et al.* *Biochemistry* **29**, 5358-5364 (1990).
- Luisi, B. F. *et al.* *Nature* **352**, 497-505 (1991).
- Schwabe, J. W., Chapman, L., Finch, J. T. & Rhodes, D. *Cell* **75**, 567-578 (1993).
- Brünger, A. T. *XPLOR Version 3.1* (Yale University Press, New Haven, 1992).
- Kurokawa, R. *et al.* *Genes Dev.* **7**, 1423-1435 (1993).
- Lee, M. S. *et al.* *Science* **260**, 1117-1121 (1993).
- Wilson, T. E., Paulsen, R. E., Padgett, K. A. & Milbrandt, J. *Science* **256**, 107-110 (1992).
- Schwabe, J. W. *et al.* *Structure* **1**, 187-204 (1993).
- Luo, Z., Rouvinen, J. & Maenpää, P. H. *Eur. J. Biochem.* **223**, 381-387 (1994).
- Madan, A. P. & DeFranco, D. B. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3588-3592 (1993).
- Schröder, M., Becker-Andre, M. & Carlberg, C. *J. biol. Chem.* **269**, 6444-6449 (1994).
- Katz, R. W. & Koenig, R. J. *J. biol. Chem.* **269**, 18915-18920 (1994).
- Wilson, T. E., Fahrner, T. J., Johnston, M. & Milbrandt, J. *Science* **252**, 1296-1300 (1991).

- Nelson, C. C., Faris, J. S., Hendy, S. C. & Romanuk, P. J. *Molec. Endocr.* **7**, 1185-1195 (1993).
- Flocco, M. M. & Mowbray, S. L. *J. molec. Biol.* **235**, 709-717 (1994).
- Xu, W., Alroy, I., Freedman, L. P. & Sigler, P. B. *Cold Spring Harbor Symp. quant. Biol.* **LVIII**, 133-139 (1993).
- Brünger, A. T. *Nature* **355**, 472-475 (1992).
- Otwinski, Z. *ML-PHARE CCP4 Proc.* 80-88 (Daresbury Laboratory, Warrington, UK, 1991).
- Jones, T. A. *Meth. Enzym.* **115**, 157-171 (1985).
- Jones, T. A. *et al.* *Acta crystallogr. A* **47**, 110-119 (1991).
- Presley, J. P. *J. appl. Crystallogr.* **21**, 572-576 (1988).
- Nicholl, A. & Honig, B. *J. J. comp. Chem.* **12**, 435-445 (1991).
- Freedman, L. *et al.* in *Proc. Ninth Workshop Vitamin D* (ed. Norman, A.) (de Gruyter, New York, in the press).
- Kurokawa, R. *et al.* *Nature* **371**, 528-531 (1994).
- Schröder, M. *et al.* *Nature* **370**, 382-386 (1994).
- Weinberger, C. *et al.* *Nature* **324**, 641-646 (1986).
- Mangelsdorf, D. J. *et al.* *Genes Dev.* **6**, 329-344 (1992).
- Gewirth, D. T. & Sigler, P. B. *Nature struct. Biol.* (in the press).
- Koelle, M. R. *et al.* *Cell* **67**, 59-77 (1991).
- De The, H., Marchio, A., Tiollais, P. & Dejean, A. *Nature* **330**, 667-670 (1987).
- Baker, A. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 3294-3298 (1988).
- Smith, D. P., Mason, C. S., Jones, E. A. & Old, R. W. *Nucleic Acids Res.* **22**, 66-71 (1994).
- Brand, N. *et al.* *Nature* **332**, 850-853 (1988).
- Krust, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 5310-5314 (1989).
- Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624-629 (1987).
- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444-450 (1987).
- Oro, A. E., McKeown, M. & Evans, R. M. *Nature* **347**, 298-301 (1990).

ACKNOWLEDGEMENTS. We thank D. Gewirth and S. Kliwer for discussion and for reagents, and S. Ealick and the staff of CHESS for access and help with the synchrotron X-ray source. Research was supported by NIH grants to P.B.S. and R.M.E. F.R. is an NIH postdoctoral fellow. T.P. was supported by a long-term fellowship from the European Molecular Biology Organization.

Evidence for a large thermal pressure imbalance in the local interstellar medium

S. Bowyer*, R. Lieu*, S. D. Sidher†, M. Lampton* & J. Knude‡

* Center for EUV Astrophysics, 2150 Kittredge Street, University of California, Berkeley, California 94720-5030, USA

† Astrophysics Group, Blackett Laboratory, Imperial College, London SW7 2BZ, UK

‡ Niels Bohr Institute for Astronomy, Physics, and Geophysics, University Observatory, Øster Voldgade 3, DK-1350 Copenhagen, Denmark

THE interstellar medium (ISM) comprises a number of components at very different densities and temperatures, from dense molecular gas at 10 K to very diffuse plasma at 10^6 K. It has generally been assumed that the warm (10^4 K) and hot ($\sim 10^6$ K) components are in thermal pressure equilibrium^{1–3}, although this has not been universally accepted⁴. Here we use the discovery of a shadow in the diffuse extreme ultraviolet background radiation⁵, cast by a nearby cloud of cool hydrogen, to calculate directly the thermal pressure in the nearby (≤ 40 pc) hot ISM. We find that it is 20 times greater than in the warm cloud surrounding the Sun. This result directly contradicts the basic assumption of thermal pressure balance in the ISM, indicating that some unidentified force (such as confinement by magnetic fields⁶) must play an important role in the overall pressure balance.

During the all-sky survey by the Extreme Ultraviolet Explorer (EUVE) satellite an extensive long exposure along the ecliptic plane was obtained by the Deep Survey (DS) telescope⁷. In this Letter we use diffuse night-time data obtained with the DS Lexan/boron filter (65–190 Å) obtained from a scan of the ecliptic plane beginning at Galactic coordinates $l \approx 156^\circ$, $b \approx -41^\circ$ and ending at $l \approx 173^\circ$, $b \approx -22^\circ$. The field of view of the DS is 1.64° in diameter, centred on the plane. The data used were from periods when the count rates did not vary with local time, solar zenith angle, observing zenith angle, or geomagnetic background; $\sim 55\%$ of the total data met these criteria. The raw count rates were corrected for background as discussed in Lieu *et al.*⁸; the diffuse astronomical flux was 25% of the total. The resulting diffuse EUV flux is shown in Fig. 1a as a function of ecliptic longitude.

The average background along the DS scan path between ecliptic longitude 39° and 60° (including the low-count-rate region centred at 52°) has a mean value of 0.487 ± 0.006 counts per second. The data were smoothed with a sliding rectangular box of 0.3° width, and reveal a region centred at 52° ecliptic longitude where the diffuse background is well below the mean⁵. The core (between 51.35° and 52.25°) and the entire region (between 50.75° and 53.25°) are respectively 3.9σ and 3.0σ below the average background.

We obtained values of the flux at $100 \mu\text{m}$ for this region, as determined by the IRAS satellite (S. Wheelock, personal communication), and averaged this data over the DS field of view. The background produced by zodiacal light and diffuse cirrus was fitted by a cubic spline function and subtracted, leaving enhanced features caused by isolated cirrus. In Fig. 1b we show the background-subtracted IRAS flux in the region of the EUV-depleted region. The enhanced flux centred at $\sim 52.5^\circ$ cannot be due to instrument effects or zodiacal light, because it is present in the same celestial position when observed by three individual IRAS scans separated in time by six months. By comparing Fig. 1a and b, it can be seen that the boundaries of the EUV-depleted region are aligned with the IRAS enhancement. There is some uncertainty associated with this alignment which is unsurprising because the hydrogen halo of an IRAS-observed cloud (which

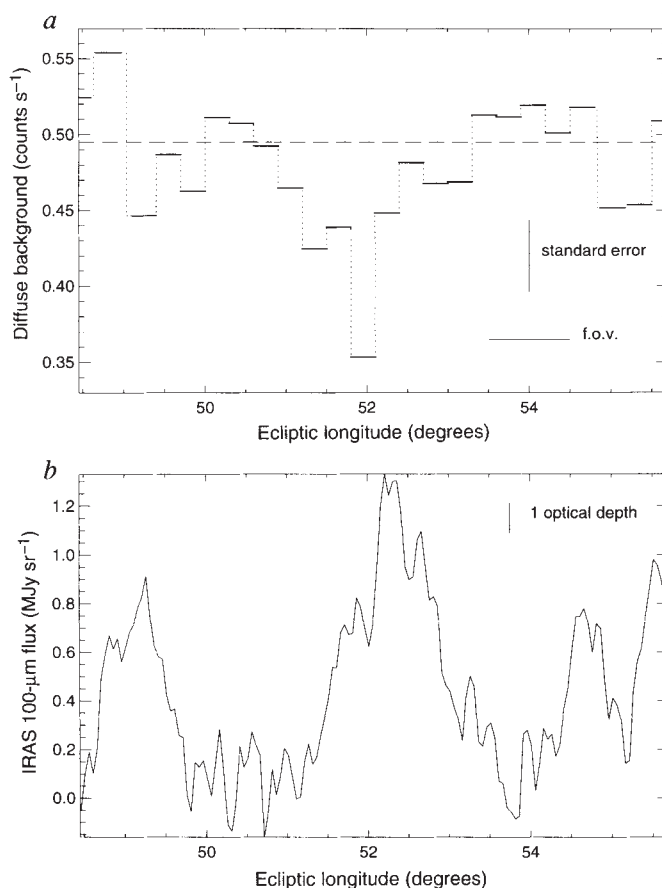


FIG. 1 a, The diffuse background (as determined by the EUVE Deep Survey (DS) telescope, with a Lexan/boron filter) as a function of ecliptic longitude (J2000) for a section of the survey scan path where a statistically significant absorption feature is found. The dashed line is the local background level, derived from data from 3.6° on each side of the feature. The 1σ error in the count rates and the exposure-averaged DS field of view (f.o.v.) are indicated. b, The IRAS $100\text{-}\mu\text{m}$ continuum-subtracted flux averaged over the DS field of view as a function of ecliptic longitude. The flux level corresponding to one optical depth of EUV absorption is marked.

would be EUV-absorbing) typically extends beyond the cloud boundary.

We can directly obtain the pressure of the hot gas in the line of sight to this cloud from the relation:

$$P = 1.92kT\sqrt{EM/L} \quad (1)$$

where T is the temperature, EM is the emission measure, L is the length of the emitting region, and k is Boltzmann's constant.

An important point is that the emitting length, L , is completely bounded because EUV emission is completely absorbed by even small amounts of gas, and (as we show in the following) no radiation comes from beyond this cloud. The mean effective wavelength for the DS data determined by a weighted integral over a nominal plasma spectrum having $T = 7 \times 10^5$ K (see below) and a foreground ISM absorption corresponding to a H I column of $N_{\text{H}} = 5 \times 10^{18} \text{ cm}^{-2}$ (to account for the local cloud around the Sun), is 126 Å . This parameter is fairly insensitive to the plasma temperature employed; a temperature as large as 3×10^6 gives a mean effective wavelength of 104 Å . We derived the mean effective wavelength with two different plasma codes^{9–11} to assess the dependency of this parameter on differing codes; the results varied by $<4\%$. The column density of hydrogen associated with a non-depleted ISM that would produce one optical depth of absorption for this mean effective wavelength is $2 \times 10^{19} \text{ cm}^{-2}$.

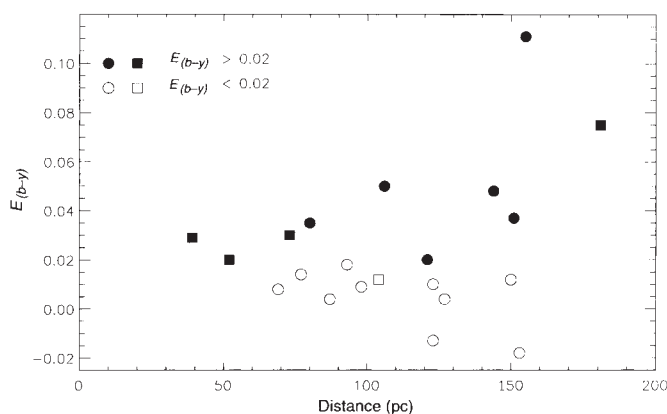


FIG. 2 Circles and squares indicate the colour excesses of stars plotted against distance. Stars with colour excesses, $E_{(b-y)} < 0.02$, are marked by open symbols, and stars with excesses of $E_{(b-y)} > 0.02$ are filled symbols. The square symbols indicate stars directly in the direction of the cloud.

The mean intensity of the IRAS cloud averaged over the 3σ EUV absorption region is $0.518 \text{ MJy sr}^{-1}$ above the baseline IRAS flux. Using the standard conversion ratio¹² $F_{\nu}(100) \approx 0.85 \text{ (MJy sr}^{-1})/10^{20} \text{ H I cm}^{-2}$, the associated hydrogen column is $6 \times 10^{19} \text{ cm}^{-2}$. The corresponding absorption, calculated using the method described in Rumph *et al.*¹³, is about three optical depths at the mean wavelength of the filter used with the DS telescope. Hence EUV flux from behind the cloud is completely blocked, and the count rate in the direction of the cloud is due entirely to foreground emission. We note that this hydrogen column corresponds to about 0.6 optical depths of absorption at the lowest energy band of the Rosat X-ray Position Sensitive Proportional Counter and so this cloud would be virtually impossible to detect with that instrument, and any emission detected from this direction would be a mixture of foreground emission and emission from beyond the cloud.

We have determined the distance to the EUV-absorbing cloud using Strömgren photometry; this technique has been shown to provide the absolute magnitude and reddening of most classes of main-sequence stars^{14,16}. High accuracy is achievable provided that the work is done with precision and that a large number of pertinent standards are observed. Photometry was carried out with the Strömgren automatic telescope at La Silla,

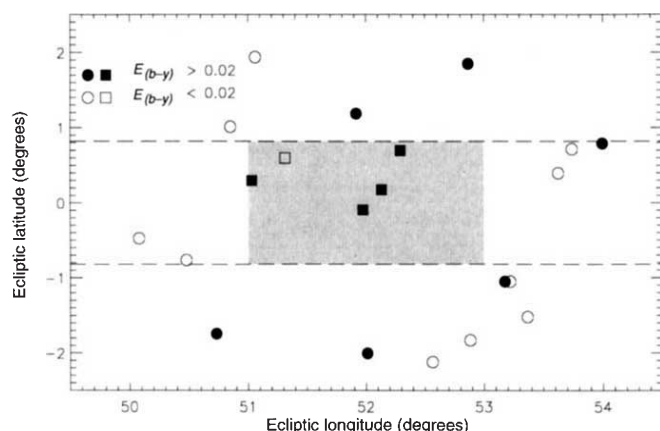


FIG. 3 Locations of stars shown in Fig. 2. The shaded area shows the position of the shadow in the diffuse EUV background observed with the EUVE DS telescope. The open and filled circles and squares are as described in Fig. 2 legend. The scan path of the EUV DS telescope is shown by dotted lines. Stars outside the cloud boundary may or may not be reddened, depending on the presence of material in these lines of sight.

Chile. We selected 91 stars for observation in a $4^\circ \times 4^\circ$ box centred on the EUV absorption feature from the PPM²¹ and the Space Telescope guide star catalogue. A number (36) of our programme stars met all criteria required for the determination of both distance and colour excess; details of this work will be reported elsewhere. Of the acceptable stars, 15 were at distances $> 200 \text{ pc}$ and provide no information of relevance to the distance to the nearby cloud. The colour excess versus distance for the remaining 21 stars in the 16 square degrees centred on the cloud are shown in Fig. 2. In Fig. 3, we show the locations of these stars in relation to the cloud shadow. Note that four of the five stars in the direction of the cloud have reddening of ≥ 0.02 , while 10 of the 16 stars outside the EUV absorption dip region have reddening of ≤ 0.02 . The single star within the (nominal) shadow with < 0.02 reddening is at the edge of the absorption region, which may indicate inhomogeneity in the cloud, or (equally likely) may be the result of measurement errors. These results show the presence of a cloud in the direction of the EUV shadow that is at a distance of $\leq 40 \text{ pc}$.

To derive the pressure with equation (1), we need the temperature of the emitting plasma; estimates of this parameter for the hot phase of the ISM vary. There is strong evidence¹⁷ for gas at $5 \times 10^5 \text{ K}$, and soft X-ray data from this hemisphere has been attributed¹⁸ to gas at a temperature of 10^6 K . The DS is most sensitive to emission from plasmas with temperature above $3 \times 10^5 \text{ K}$ and is quite insensitive to plasmas with temperatures $> 2 \times 10^6$ (ref. 7); Lieu *et al.*⁸ find a temperature of $7 \times 10^7 \text{ K}$ for the diffuse flux seen by EUVE. To first order we assume a single-temperature plasma at $7 \times 10^5 \text{ K}$, but we note that our final result is not strongly dependent on this assumption.

We derive the emission measure for the plasma in the line of sight to the cloud by combining the DS count rate in the direction of the cloud, 0.437 ± 0.02 counts per second, with a plasma emissivity code¹¹ for a temperature of $7 \times 10^5 \text{ K}$. The emission measure is $0.0077 \text{ cm}^{-6} \text{ pc}$. Using this value in equation (1), we obtain a pressure (P/k) of $19,000 \text{ cm}^{-3} \text{ K}$.

We can now test the assumption of thermal pressure equilibrium by comparing the pressure of the local hot phase of the ISM obtained here with the pressure of the warm local interstellar cloud surrounding the Sun. The pressure of the local cloud has been well established by a variety of techniques, the most direct and unambiguous of which is to analyse solar He I 584 Å radiation resonantly scattered by helium in the inflowing cloud. The pressure obtained¹⁹ is $730 \pm 30 \text{ cm}^{-3} \text{ K}$.

The wide difference (a factor of almost 30) between these two pressures directly challenges ISM models in which thermal pressure equilibrium is a basic assumption. In the McKee-Ostriker model³ (which falls in this class) the pressure at any particular location is established by the effects of past supernovae, and disequilibrium can exist for a brief period after a supernova explosion. No recent supernova has affected our Local Region; Frisch¹⁹ and Bertin *et al.*²⁰ have provided independent evidence that this region has not been violently heated within at least the past million years.

We have tried to imagine circumstances under which our data might be compatible with a lower pressure. The pressure obtained if the cloud is closer than 40 pc , or if the ISM abundances are depleted, or if the hot material is clumped, is always higher than the nominal value derived above. The only circumstances which could lead to a lower pressure is the existence of an additional unknown source of background which has uniformly corrupted our data. In this case, the minimum pressure obtainable without producing other gross effects (for example, requiring the EUV background from the non-occluded section of the scan to be produced at distances $> 200 \text{ pc}$) is $> 7,000 \text{ cm}^{-3} \text{ K}$. The result produced in even this unlikely case will not affect our conclusions.

If the warm and hot phases of the ISM are not in thermal pressure equilibrium, why do cooler clouds persist? This is certainly a question for further investigation; an attractive alterna-

tive is the suggestion by Cox and co-workers⁶ that confinement is by magnetic fields. □

Received 11 November 1994; accepted 27 March 1995.

1. Spitzer, L. *Astrophys. J.* **124**, 20–34 (1956).
2. Field, G., Goldsmith, D. W. & Habing, H. *Astrophys. J.* **155**, L149–L154 (1969).
3. McKee, C. & Ostriker, J. *Astrophys. J.* **218**, 148–169 (1977).
4. Slavin, J. & Cox, D. *Astrophys. J.* **417**, 187–195 (1993).
5. Lieu, R., Sumner, T. J., Bowyer, S. & Sidher, S. D. *Bull. Am. astr. Soc.* **25**, 860 (1993).
6. Cox, D. P. & Snowden, S. L. *Adv. Space Res.* **6**, 97–107 (1986).
7. Bowyer, S. & Malina, R. F. in *EUV Astronomy* (eds Malina, R. F. & Bowyer, S.) 397–408 (Pergamon, New York, 1991).
8. Lieu, R., Bowyer, S., Lampton, M., Jelinsky, P. & Edelstein, J. *Astrophys. J.* **417**, L41–L44 (1993).
9. Mewe, R., Gronenschild, E. H. B. M. & van den Oord, G. H. J. *Astr. Astrophys. Supp.* **62**, 197–254 (1985).
10. Mewe, R., Lemen, J. R. & van den Oord, G. H. J. *Astr. Astrophys. Supp.* **65**, 511–536 (1986).
11. Landini, M. & Monsignori-Fossi, B. C. *Astr. Astrophys. Supp.* **82**, 229–260 (1990).
12. Boulanger, F. & Perault, M. *Astrophys. J.* **330**, 964–985 (1988).
13. Rumph, T., Bowyer, S. & Vennes, S. *Astr. J.* **107**, 2108–2114 (1994).
14. Strömberg, B. A. *Rev. Astr. Astrophys.* **4**, 433–472 (1966).
15. Crawford, D. L. *Astr. J.* **84**, 1858–1865 (1979).
16. Olsen, E. H. *Astr. Astrophys.* **189**, 173–178 (1988).
17. Martin, C. & Bowyer, S. *Astrophys. J.* **350**, 242–261 (1990).
18. Snowden, S., Schmitt, J. & Edwards, B. *Astrophys. J.* **364**, 118–122 (1990).
19. Frisch, P. *Science* **265**, 1423–1427 (1994).
20. Bertin, P., Lallemand, R., Ferlet, R. & Vidal-Madjar, A. *Astr. Astrophys.* **278**, 549–560 (1993).
21. Roeser, S. & Bastien, U. *Position and Proper Motion Star Catalogue* (Spektrum, Heidelberg, 1991).

ACKNOWLEDGEMENTS. We thank G. Field and C. McKee for useful discussions. S.B., R.L. and M.L. were supported by NASA; J. Knude thanks the European Southern Observatory for observing time and travel support.

Imaging with an atomic beam

W. G. Kaenders*, F. Lison*†, A. Richter*,
R. Wynands*† & D. Meschede*†

* Institut für Quantenoptik, Universität Hannover, Welfengarten 1,
D-30167 Hannover, Germany

FOCUSING of atomic beams has been investigated for more than 40 years^{1–9}. The formation of images of simple objects such as points or slits has been demonstrated^{7–9}, but only with long focal lengths and large chromatic aberrations owing to the velocity distribution of the thermal beam. Lasers can be used to slow and cool the atomic beam¹⁰, reducing the focal length; and by optically narrowing the velocity spread of the beam, chromatic aberrations can be reduced substantially. Here we report the construction of an atomic imaging device, using a hexapole lens made from permanent magnets, which produces images in the same way as an optical slide projector. A mask is illuminated by a beam of caesium atoms prepared using a laser diode, and its image appears on a 'resonant light screen', where a sheet of laser light excites atomic fluorescence. We can obtain magnified and demagnified images with remarkable resolution, suggesting that this technique might be used to create sub-micrometre atomic structures.

In our experiment, a caesium atomic beam is laser-slowed, cooled, and polarized by the light chirp technique^{11,12} in a vacuum chamber of 1,285 mm length before it arrives at the imaging zone (Fig. 1). At the end of the laser cooling cycle, 95% of the atoms are optically pumped into the extreme magnetic substate ($F=4$, $m_F=+4$) of the $6^2S_{1/2} \rightarrow 6^2P_{3/2}$ transition at 852 nm wavelength and have residual velocities in the range 150 m s^{-1} to zero, depending on when the frequency chirp is stopped. The experimental uncertainty of the final velocity is $\pm 3 \text{ m s}^{-1}$. However, the final velocity spread due to the cooling itself is smaller, of the order of $\Delta v_z \approx 1 \text{ m s}^{-1}$ where v_z is the velocity component along the beam axis. A mechanical shutter blocks the atomic oven when the cooling cycle has started, to prevent thermal

atoms ($v_z = 300 \text{ m s}^{-1}$) that leave the oven too late to take part in the cooling process from arriving in the imaging zone at the same time as the cooled atoms.

The slowing and cooling process is finished, and the preparing lasers are shut off, by the time the atoms reach the mask. There the cold atomic pulse has a transverse diameter of 7 mm (full-width at half-maximum) so that a fairly large mask can be used. In our case it consists of a thin transparent piece of poly(methyl methacrylate) with holes drilled through it. Behind the mask, the transmitted atoms fly freely towards the magnetic hexapole lens (object distance g), consisting of 12 homogeneously magnetized segments made from NdFeB permanent magnetic material¹³. The magnetization direction for each segment is oriented such that a plane magnetic hexapole is formed¹⁴ (Fig. 2). In the inner bore, the magnetic flux density increases quadratically in the radial (ρ) direction. Therefore the atoms experience a gradient force linear in ρ due to the magnetic potential, making the hexapole a complete analogue of the optical gradient-index lens. Within a thin-lens approximation, that is, for a short hexapole, the focal length is a quadratic function of atomic velocity.

The atomic image is detected at a distance b downstream from the lens. There the atoms traverse a sheet of resonant laser light and the resulting fluorescence is detected by a CCD (charge-coupled device) camera¹⁵ (Fig. 1).

Sequences of video frames for $g=98 \text{ mm}$, $b=120$ or 80 mm , and several values of final atomic velocity v_z (and thus focal lengths of the lens) were taken. The camera was switched on only when the image-forming atoms were passing through the detection zone. Because the fast atoms had already passed by then they do not appear in the video frames (Fig. 3A, B).

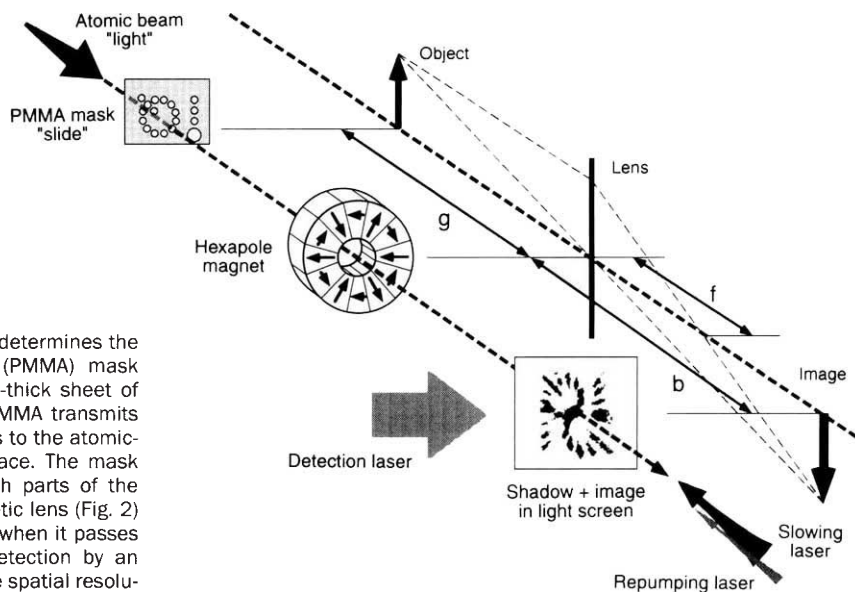
The atomic distribution in the light screen changes dramatically with final velocity v_z . With an imaging distance of $b=120 \text{ mm}$, a minimum spot size is obtained for $v_z=110 \text{ m s}^{-1}$. Focal spot diameters of the order of the spatial resolution of the detection scheme ($200 \mu\text{m}$) have been recorded in single-shot experiments without the mask in place, giving an upper limit of 1 m s^{-1} for the longitudinal velocity scatter of the beam. The focal spot has a clean gaussian density distribution, as expected for the lens image of a gaussian beam with an increase in density by a factor of 800 in comparison to the thermal beam. For smaller velocities the evolution of an inverted image can be seen. A final velocity of $v_z=70 \text{ m s}^{-1}$ produces a sharp image (Fig. 3B). It has a magnification of $b/g \approx 1.4$. For further decreased atomic velocity v_z the image becomes more and more defocused. Below $v_z=50 \text{ m s}^{-1}$ the contrast is too low for detection of an image. The same experiment was repeated for a shorter imaging distance $b=80 \text{ mm}$, resulting in a sharp image for a velocity of $v_z=60 \text{ m s}^{-1}$ and a magnification of $b/g \approx 0.8$ (Fig. 3A). The dashed outlines indicate the shape of the original mask, properly scaled with the ratio b/g . With a wider time gate for the camera, atoms in the thermal part of the beam also show up in the video frames. Because they are too fast to experience any substantial deflecting force, they form a 'shadow' of the object in the detection zone (Fig. 3C). It should be noted that the mask was mounted off-axis in order to be able to distinguish the shadow and the inverted image.

For background-free imaging, the slowed part of the atomic beam should be spatially separated from the remaining thermal beam. We have done this using an off-axis atom optical telescope constructed from two magnetic lenses. This will also remove the technique's restriction to masks which are transparent to the slowing laser light.

Because of the off-axis illumination, lens aberrations are more pronounced in the outer regions of the image. In principle, the lens suffers from all the geometric aberrations known from light optics. Here longitudinal and transversal chromatic aberrations are particularly important. In addition, the image is also affected by imperfect atomic-beam collimation and polarization, deviations from a pure magnetic hexapole, gravity and atomic collisions.

† Present address: Institut für Angewandte Physik, Universität Bonn, Wegelerstrasse 8, D-53115 Bonn, Germany.

FIG. 1. Diagram of apparatus, showing the caesium atomic beam with the slowing and repumping laser and the imaging zone (object, lens, screen). The caesium beam is collimated to a half angle of $1/360$ rad. A small axial magnetic field (5 G) is applied along the atomic beam line. The Doppler shift brings the atoms into resonance with the red-detuned slowing laser beam. A net force is exerted on the atoms because they absorb photons from the laser beam's direction only but reradiate them isotropically by spontaneous emission. When the atoms become slower, their Doppler shift decreases and is compensated by a synchronous chirp of the laser frequency¹¹. The final laser frequency determines the final atomic velocity. The poly(methyl methacrylate) (PMMA) mask ('slide' or object of the projection) consists of a 1-mm-thick sheet of PMMA centred slightly above the atomic beam axis. PMMA transmits 90% of the incident laser light, and no disturbing effects to the atomic-beam preparation are observable with the mask in place. The mask has 18 drilled holes (diameter 500 μm) through which parts of the atomic beam can pass. At a fixed distance g the magnetic lens (Fig. 2) is mounted. A detection laser excites the atomic beam when it passes through a sheet of resonant light for fluorescence detection by an image-intensified and gateable CCD video camera¹⁵. The spatial resolution is typically 200 μm . On-line read-out of the camera (up to two images per second) can be used to align the imaging components. A typical time for recording video frames with a contrast of ~ 100 is 3 s (30 slowing/imaging cycles).

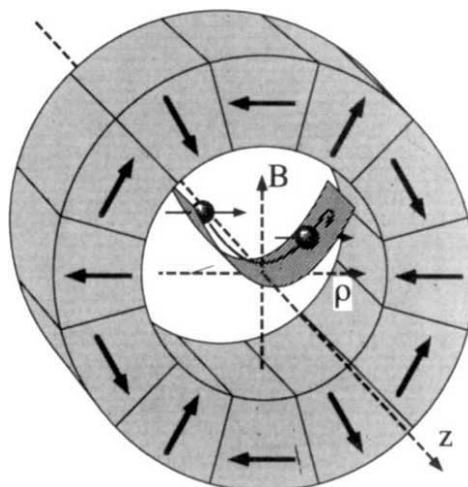


To understand the influence of the various aberrations on the minimum focal spot size, we have numerically simulated atomic trajectories in a realistic thick lens and compared them to analytical approaches similar to those described by McClelland *et al.*⁴. This work will be reported in detail elsewhere. The field generated by the segmented permanent-magnet assembly can be described analytically in good approximation (including the fringe field and the deviations from the ideal plane hexapole), thus allowing mathematical treatment of aberrations in the image similar to methods used in designing electron optics⁴. Most aberrations can be minimized by reducing the initial beam diameter, by blocking the inner part of the lens (which is most sensitive to small stray fields), and by improving the collimation and alignment of the atomic beam. Optical preparation of the atomic beam strongly reduces the chromatic aberration. At present, the minimum achievable longitudinal velocity spread is limited by our cooling technique to $\sim v_z/\Delta v_z = 100$ which limits the spot size to $\sim 0.5 \mu\text{m}$ for

$v_z \approx 50 \text{ m s}^{-1}$ and a beam diameter of 10 mm. If the beam diameter is reduced to 1 mm, chromatic aberration (for $v_z/\Delta v_z = 100$) allows a spot size of 50 nm to be achieved. Our numerical calculations that take into account all other classical aberrations also give 50 nm spot size. This resolution is clearly of potential interest for lithographic applications. The diffraction limit for the atomic de Broglie wave for a lens with an aperture diameter of 1 mm and caesium atoms with $v_z = 70 \text{ m s}^{-1}$ is 6.5 nm. It is possible to explore this regime using advanced cooling schemes at the so-called recoil limit¹⁶ and transverse cooling with optical molasses^{17,18}.

This imaging technique has interesting potential applications in nanostructure fabrication. When the detection light screen is replaced by a substrate, the structure of an arbitrarily complex mask—which might also be a standing light field or even a holographic light pattern—can be translated into a pattern of deposited atoms⁶. The deposition occurs fully in parallel over a millimetre-sized region with high atomic flux, and with an imag-

FIG. 2. Diagram of the hexapole lens. The magnetically extremely hard material (NdFeB) used for the lens has a remanence of 1.1 T and a coercive force of $1,670 \text{ kA m}^{-1}$. The lens has an outer diameter of $d_a = 50 \text{ mm}$, an open bore of $d_i = 15 \text{ mm}$, and a geometric length of $L = 50 \text{ mm}$. The magnetic flux density increases quadratically with ρ , as shown schematically inside the bore. The maximum radial curvature of the magnetic flux density B near its symmetry axis in the centre of the hexapole is $\partial^2 B/\partial \rho^2 = 2.66 \text{ T cm}^{-2}$. Inside the inner bore of the hexapole magnet the atoms experience a gradient force $\mathbf{F}$ due to the conservative magnetic potential $\mathbf{F} = -\nabla(\mu_B \cdot \mathbf{B})$ where μ_B is the Bohr magneton. For the polarized atomic beam in the upper extreme (F, m_F) substate, the coupling of the atom to the magnetic field equals μ_B independent of magnetic field strength. Therefore the force directed towards the axis is linearly dependent on the radial coordinate ρ , as required for a lens: $\mathbf{F} = -\mu_B \nabla |\mathbf{B}| \approx -\mu_B \rho \partial^2 B/\partial \rho^2 \mathbf{e}_\rho$. A parallel atomic beam acquires a radial velocity component which is inversely proportional to the longitudinal atomic velocity, leading to oscillations of the atoms about the z -axis for long lenses or very slow atoms. Within the thin-lens approximation, the focal length is a quadratic function of atomic velocity $f = m v_z^2 / (2\mu_B \int_0^L \partial^2 B/\partial \rho^2(z) dz)$. The imaging distance b can then be found (to a good approximation) using the well-known thin lens equation, where in our experiments the object distance $g = 98 \text{ mm}$ is kept constant and f can be tuned by varying the final velocity v_z . The uncertainty of the atomic velocity converts to a typical error for the focal length of a few millimetres.



ing ratio of 1:100 the deposition of nanostructures should be possible using micro-structured masks.

The magnetic imaging technique introduced here is generally applicable. Because magnetic lenses operate in the same manner on all atomic species possessing a magnetic dipole moment, a large number of chemical elements can be manipulated. The light forces, requiring very specific laser equipment, are used only for the optical preparation of the beam: slowing, reducing the velocity spread (cooling), and optically pumping all atoms into the same magnetic substate (polarization). For several technologically important materials (for example, Ag, Al and Cr) laser cooling has already been demonstrated, so our technique should find practical applications, particularly when used in combina-

tion with a high-brightness atomic beam source, such as a Zeeman slower^{19,20}. □

Received 3 October 1994; accepted 11 April 1995.

- Adams, C. S., Sigel, M. & Mlynek, J. *Phys. Rep.* **240**, 143–210 (1994).
- Friedburg, H. & Paul, W. *Naturwissenschaften* **38**, 159–160 (1951).
- Friedburg, H. Z. *Phys.* **130**, 493–512 (1951).
- McClelland, J. & Scheinfein, M. R. *J. opt. Soc. Am.* **B8**, 1974–1986 (1991).
- McClelland, J. J., Scholten, R. E., Palm, E. C. & Celotta, R. J. *Science* **162**, 877–880 (1993).
- Timp, G. et al. *Phys. Rev. Lett.* **69**, 1636–1639 (1992).
- Sleator, T., Pfau, T., Balykin, V. & Mlynek, J. *Appl. Phys.* **B54**, 375–379 (1992).
- Balykin, V. I., Letokhov, V. S., Sidorov, A. I. & Ovchinnikov, Y. B. *J. mod. Opt.* **35**, 17–34 (1988).
- Carnal, O., Sigel, M., Sleator, T., Takuma, H. & Mlynek, J. *Phys. Rev. Lett.* **67**, 3231–3234 (1991).
- J. opt. Soc. Am. (Spec. Iss.)* **B6**, (1989).
- Ertmer, W., Blatt, R., Hall, J. L. & Zhu, M. *Phys. Rev. Lett.* **54**, 996–999 (1985).
- Watts, R. N. & Wieman, C. E. *Opt. Lett.* **11**, 291–293 (1986).
- Kaenders, W. G., Frerichs, V., Schröder, F. & Meschede, D. *Hyperfine Interactions* **76**, 221–232 (1993).
- Halbach, K. *Nucl. Instrum. Meth.* **169**, 1–10 (1980).
- Esslinger, T., Hemmerich, A. & Hansch, T. W. *Opt. Commun.* **93**, 49–53 (1992).
- Kasevich, M. et al. *Phys. Rev. Lett.* **66**, 2297–2301 (1991).
- Nellessen, J., Werner, J. & Ertmer, W. *Opt. Commun.* **78**, 300–308 (1990).
- Witte, A., Kisters, T., Riehle, F. & Helmcke, J. *J. opt. Soc. Am.* **B9**, 1030–1037 (1992).
- Prodan, J. V., Phillips, W. D. & Metcalf, H. *Phys. Rev. Lett.* **49**, 1149–1153 (1982).
- Joffe, M. A., Ketterle, W., Martin, A. & Pritchard, D. E. *J. opt. Soc. Am.* **B12**, 2257–2262 (1992).

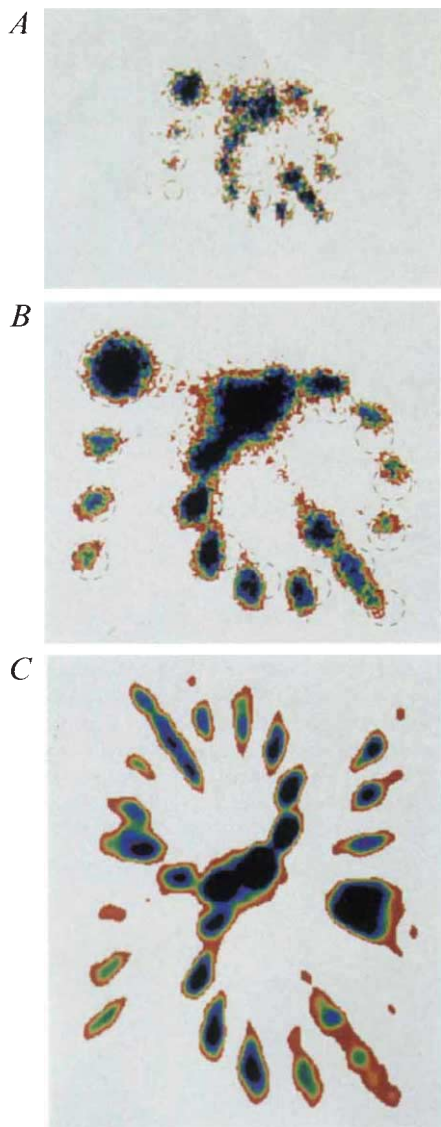


FIG. 3 Video frames showing sharp images at two different focal lengths and imaging distances (A, $b=80$ mm; B, $b=120$ mm). The magnification factors are 0.8 and 1.4, respectively. For comparison the shape of a perfect image is indicated by dashed lines. The bright spot in the fluorescence video frames is caused by atoms which have escaped the slowing procedure, and thus marks the optic axis. Because these atoms are faster than the rest they appear closer to the focal spot of the lens than desired and smear out the image. When not only laser-cooled but also thermal atoms arrive in the detection zone while the camera is switched on, both the shadow produced by the undeflected fast atoms and the image can be seen in the same video frame, clearly demonstrating the formation of an inverted real image (C).

Structural quantum effects and three-centre two-electron bonding in CH_5^+

Dominik Marx*† & Michele Parrinello†

* IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

† Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

HYPERCOORDINATE carbonium ions can be formed by protonating saturated hydrocarbons with superacids^{1–3}. As this leaves a deficiency of bonding electrons, the resulting non-classical carbocations contain bonds in which two electrons are shared between three nuclei. Protonated methane, CH_5^+ , might be seen as the prototype of such species^{1–3}. But recent calculations^{4,5} have suggested that all five C–H bonds are effectively equivalent and exchange dynamically very rapidly. It was therefore concluded⁴ that CH_5^+ is a highly fluxional molecule without a definite structure, in which the representation in terms of three-centre two-electron bonding is misleading. Here we use a recently developed technique⁶ to perform *ab initio* electronic structure calculations that include quantum effects of the nuclei. We find that, although there are prominent quantum-mechanical effects on the structure, including fluxionality, pseudo-rotations and hydrogen scrambling, the quantum ground state is nevertheless dominated on average by configurations in which an H_2 moiety is attached to a CH_3 group forming a three-centre two-electron bond. To this extent, CH_5^+ should therefore resemble other carbonium ions.

According to the standard picture of bonding in carbonium ions^{1–3}, the parent CH_5^+ is believed to have an eclipsed ground state of point symmetry C_s , in which two paired protons constituting an H_2 moiety are attached to a CH_3 tripod. This results in three two-centre two-electron (2c–2e) bonds and one 3c–2e bond^{1–3}. There is essentially no energy barrier to the rotation of the H_2 moiety around the pseudo C_3 -axis of the CH_3 group to a staggered C_s first-order saddle point. Characteristic of both C_s structures are three short C–H bonds in the CH_3 tripod, two long C–H bonds to the H_2 moiety, and an elongated H–H bond.

† Present address: Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany.

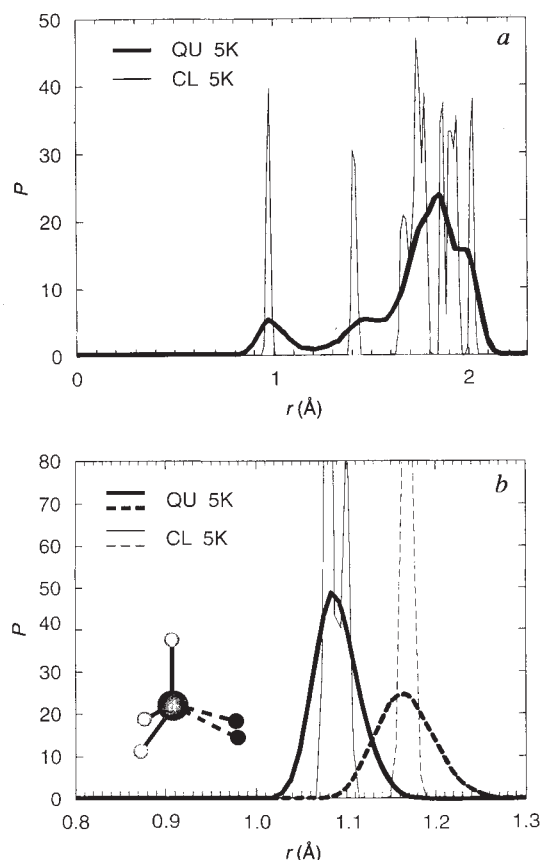


FIG. 1 Radial distribution functions: *a*, of the H-H distances in CH_5^+ (normalized to the total number of H-H bonds); *b*, of the C-H bond lengths in CH_5^+ , where bond lengths involving the protons engaged in the H_2 moiety are shown by dashed lines, and bond lengths involving the other protons are shown by solid lines (see inset). The peak around 1 Å in *a* stems from a closely bound pair of protons, that is, from an H_2 moiety, and decays to a minimum near 1.2 Å. In addition it integrates to unity, corresponding to a single H-H bond at the same value. Based on this, we used a cut-off H-H bond length of 1.15 Å to define the H_2 moiety, and found that our conclusions are stable within reasonable variations. Bold lines indicate the results of quantum *ab initio* path integral simulations at 5 K; note that these are quantum-thermal averages in the sense of quantum statistical mechanics, which correspond at the chosen temperature to the quantum ground state. Light lines indicate results obtained with the classical approximation to the nuclei at the same average temperature. The underlying electronic structure calculations are based on Becke's exchange-gradient-corrected local density functional method¹⁸. The core electrons are represented by soft non-local von Barth-Car type pseudopotentials. The valence orbitals are expanded in a plane-wave basis up to 35 Rydberg without imposing periodic boundary conditions on the simple cubic supercell of length 20 atomic units (1 a.u. = 0.529 Å). The mapping onto $P=32$ replicas in imaginary time is sufficient. The quantum computations were performed in parallel on the IBM scalable POWERparallel System 9076 SP1.

The next low-lying saddle point with C_{2v} point-group symmetry does not possess such an H_2 moiety; see the figure in ref. 5 for these structures.

Direct experiments^{5,7-9,19} to determine the structure of CH_5^+ unambiguously have been unsuccessful. But sophisticated calculations have revealed extremely small energy differences between C_s and C_{2v} structures^{4,10,11} so that zero-point motion quantum effects might play a decisive role, an idea taken to the extreme in the proposal^{4,5} that CH_5^+ is a molecule without a definite structure.

To shed light on this puzzle of both fundamental and practical relevance, we have used a technique recently introduced by us⁶. In the spirit of the Car-Parrinello method¹²⁻¹⁵, the electronic structure is calculated from density functional theory as the

nuclei are propagated. The crucial extension, however, is the inclusion of quantum dispersion effects of the nuclei in the well established framework of Feynman's path-integral formulation of quantum statistical mechanics¹⁶. Following this approach, the six quantum nuclei of CH_5^+ are mapped onto P cyclically connected replicas which interact according to a suitably scaled Born-Oppenheimer energy surface; the mapping becomes exact as $P \rightarrow \infty$. This mapping is often described as a cyclical evolution of the molecule in imaginary time¹⁷. The fluctuations of the replicas along the path provide, in terms of imaginary time correlation functions, a rigorous measure of quantum dispersion effects; the replicas become identical in the classical limit. This formulation allows us to sample, in the sense of quantum statistical mechanics, the molecule's finite-temperature many-body nuclear density matrix as well as operator expectation values based on an *ab initio* electronic ground-state surface generated as configuration space is explored. In this manner, quantum and thermal excitations (for example, zero-point vibrations) are rigorously included as governed by the proper quantum statistical distribution without further approximations, such as the harmonic approximation which would be invalid for a highly fluxional molecule like CH_5^+ . A standard Car-Parrinello molecular dynamics study of CH_5^+ at room temperature, where the nuclei are treated in the classical approximation, is reported in ref. 20.

By comparing the quantum to the classical curves (Fig. 1), we find pronounced structural quantum effects on radial distribution functions as hypothesized in refs 1, 2. These effects are in principle measurable by gas-phase electron diffraction. The ground-state distribution of H-H distances (Fig. 1*a*) is significantly broadened by quantum fluctuations, which allow H-H separations between those of the peak at small distances and the others. This indicates that there is tunnelling between various structures, because the peak at an H-H separation of $r_{\text{HH}} \approx 1$ Å clearly stems from the H_2 moiety whereas the contributions for $r_{\text{HH}} \geq 1.4$ Å are due to the other non-bonded H-H interactions. However, the existence of a single well separated peak from the H_2 moiety, as well as the preservation of the general shape of the distribution, already suggest that CH_5^+ may still have a relatively well defined ground-state structure.

Valuable structural insight can be obtained by inspecting molecular configurations sampled along the cyclic path in imaginary time¹⁷. In Fig. 2*a-c*, local fluctuations lead to the rotation of the H_2 moiety and interconvert eclipsed and staggered C_s -like structures. However, CH_5^+ finally pseudorotates to the structure shown in Fig. 2*d* by changing the identity of the protons constituting the H_2 moiety. The crucial point, however, is that the chemical structure of CH_5^+ characterized by an H_2 moiety, which is 3c-2e bonded to the CH_3 tripod shown in Fig. 2*a-d*, is essentially preserved during this hydrogen scrambling. Moreover, structures without an H_2 moiety (Fig. 2*e*) play only a minor role. The superposition of all P configurations that complete a single trajectory path in imaginary time (Fig. 2*f*) demonstrates that successive pseudorotations lead in effect to a large-amplitude 'floating' of an H_2 moiety around the carbon atom, while preserving the molecule's 3c-2e bonded C_s -like structure.

Let us now quantify these findings. Configurations such as those in Fig. 2*a-d* that possess an H_2 moiety contribute to the quantum ground state with a weight of >80%. We now show that their structural properties are very close to those of the classical C_s ground state. First of all, we separate in Fig. 1*b* the C-H bond-length distribution of the protons engaged in the H_2 moiety (dashed lines) from that of the others. Despite broadening, we find also in the quantum case a clear C-H bond-length separation (mean and r.m.s. widths: 1.166 ± 0.03 Å, 1.088 ± 0.02 Å) that is very similar to the classical distribution (1.168 ± 0.01 Å, 1.086 ± 0.01 Å), and thus to the classical C_s ground state. The latter is usually described in terms of 3c-2e bonding of the H_2 moiety to the CH_3 tripod¹⁻³. Further structural information can be obtained from the distribution of the

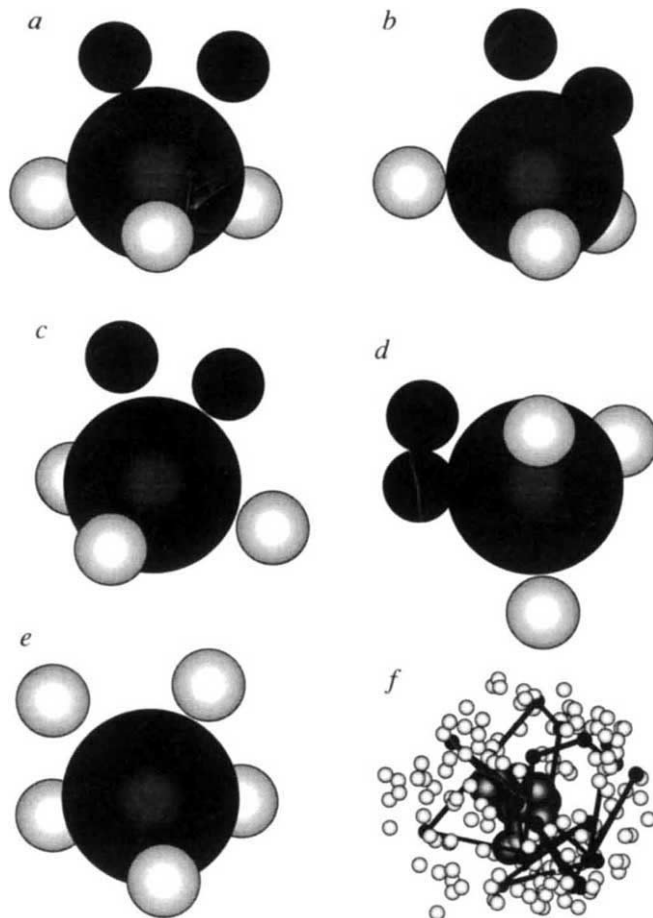


FIG. 2 *a–d*, Typical sequence in imaginary time¹⁷ of configurations of CH_5^+ with H_2 moiety (protons constituting the H_2 moiety are shown in black) *e*, structure without H_2 moiety. *f*, Full imaginary time path representing a “configurational snapshot”, which contributes to the nuclear density¹⁷; here the centres of mass of the H_2 moiety are shown in black and are connected sequentially by lines. *a–c* depict how local quantum fluctuations lead only to a rotation of the H_2 moiety, and thus to a conversion between eclipsed and staggered C_s -like structures. In evolving to *d*, however, the molecule exchanges the protons constituting the H_2 moiety, and thus pseudorotates to a configuration which has again a C_s -like structure. The average H–H bond length of the H_2 moiety and its r.m.s. fluctuations are $1.00 \pm 0.07 \text{ \AA}$ and $0.978 \pm 0.01 \text{ \AA}$ in the quantum and classical cases, respectively. Structures without any H_2 moiety such as that in *e* also occur, but they contribute only little to the average structural characteristics of CH_5^+ . *f* illustrates the large-amplitude motion due to pseudorotations, and the resulting fluxionality including hydrogen scrambling. Note also the significant quantum dispersion effects on the carbon atom, which we found to be important for the fluxionality of the CH_5^+ structure. With our methodology for the underlying electronic structure calculations (see Fig. 1 legend), we obtain for CH_5^+ —in the nomenclature of ref. 4—the following bond lengths ($\text{\AA}$) for the eclipsed $\text{C}_s(\text{I})$ classical ground state $r_{\text{CH}_3} = 1.103$, $r_{\text{CH}_6} = 1.174$, $r_{\text{CH}_7} = 1.169$, $r_{\text{CH}_8} = 1.084$, $r_{\text{H}_2\text{H}_6} = 0.978$, and the C_{2v} first-order saddle point $r_{\text{CH}_3} = 1.148$, $r_{\text{CH}_6} = 1.127$, $r_{\text{CH}_7} = 1.083$ with a classical energy difference of $0.6 \text{ kcal mol}^{-1}$.

angle formed by the vectors pointing from the carbon atom to the centre of mass of the H_2 moiety and to the centre of mass of the non-bonded protons. In the quantum case it has a peak at $162 \pm 7^\circ$, which should be compared to the classical and the ground-state values of $164 \pm 1^\circ$ and 164.2° , respectively. Again, peak position and sharpness confirm the notion of a well defined C_s -like structure. Clearly, this cannot be the whole story. The hydrogen scrambling shown in Fig. 2*f* is quantified based on imaginary time correlation functions¹⁷; amplitude and angle of the pseudorotations are $\sim 1.4 \text{ \AA}$ and $\sim 80^\circ$, respectively. Thus,

CH_5^+ undergoes large-amplitude pseudorotations, which result in hydrogen scrambling and statistically equivalent protons. Nevertheless there is an overwhelming probability of finding CH_5^+ in a structure similar to the classical C_s ground state.

Thus, our calculations suggest that CH_5^+ is indeed a highly fluxional molecule undergoing pseudorotations and hydrogen scrambling, as concluded also in refs 4 and 5. This rather complex dynamics makes the spectroscopic properties hard to determine. But the presence of an H_2 moiety is a characteristic feature of its quantum ground state, which can best be pictured as being C_s -like, with $3c-2e$ bonding. Thus, pronounced tunnelling does not destroy the chemical structure of the classical ground state, and in this sense CH_5^+ has strong similarities with typical carbenium ions^{1,3}. □

Received 14 November 1994; accepted 12 April 1995.

1. Olah, G. A. *Carbocations and Electrophilic Reactions* (Chemie, Weinheim, 1974).
2. Olah, G. A., Prakash, G. K. S., Williams, R. E., Field, L. D. & Wade, K. *Hypercarbon Chemistry* (Wiley, New York, 1987).
3. Olah, G. A., Prakash, G. K. S. & Sommer, J. *Superacids* (Wiley, New York, 1985).
4. Schreiner, P. R., Kim, S.-J., Schaefer, H. F. & Schleyer, P. v. R. *J. chem. Phys.* **99**, 3716–3720 (1993).
5. Scuseria, G. A. *Nature*, **366**, 512–513 (1993).
6. Marx, D. & Parrinello, M. *Z. Phys. B (Rapid Note)* **95**, 143–144 (1994).
7. Boo, D. W. & Lee, Y. T. *Chem. Phys. Lett.* **211**, 358–363 (1993).
8. Hiraoka, K., Kudaka, I. & Yamabe, S. *Chem. Phys. Lett.* **184**, 271–276 (1991).
9. Heck, A. J. R., de Koning, L. J. & Nibbering, N. M. M. *J. Am. Soc. Mass Spectrom.* **2**, 453–458 (1991).
10. Kloppe, W. & Kutzelnigg, W. *J. phys. Chem.* **94**, 5625–5630 (1990).
11. Schleyer, P. v. R. & Carneiro, J. W. d. M. *J. comput. Chem.* **13**, 997–1003 (1992).
12. Car, R. & Parrinello, M. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
13. Jones, R. O. & Gunnarsson, O. *Rev. mod. Phys.* **61**, 689–746 (1989).
14. Remler, D. K. & Madden, P. A. *Molec. Phys.* **70**, 921–966 (1990).
15. Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. *Rev. mod. Phys.* **64**, 1045–1097 (1992).
16. Feynman, R. P. & Hibbs, A. R. *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
17. Chandler, D. in *Liquids, Freezing, and Glass Transition* (eds Hansen, J. P., Levesque, D. & Zinn-Justin, J.) 193–285 (Elsevier, Amsterdam, 1991).
18. Becke, A. D. *Phys. Rev. A* **38**, 3098–3100 (1988).
19. Boo, D. W., Liu, Z. F., Tse, J. S., Suits, A. G. & Lee, Y. T. *Phys. Rev. Lett.* (submitted).
20. Tse, J. S., Klug, D. D. & Laasonen, K. *Phys. Rev. Lett.* **74**, 876–879 (1995).

ACKNOWLEDGEMENTS. D.M. acknowledges the Forschungsspendium from the Deutsche Forschungsgemeinschaft and an IBM postdoctoral research fellowship.

Spectroscopic evidence against nitric acid trihydrate in polar stratospheric clouds

Owen B. Toon* & Margaret A. Tolbert†

* NASA Ames Research Center, Moffett Field, California 94035, USA

† Department of Chemistry and Biochemistry and CIRES, University of Colorado, Boulder, Colorado 80309, USA

HETEROGENEOUS reactions on polar stratospheric clouds (PSCs) play a key role in the photochemical mechanism thought to be responsible for ozone depletion in the Antarctic and the Arctic^{1,2}. Reactions on PSC particles activate chlorine to forms that are capable of photochemical ozone destruction, and sequester nitrogen oxides (NO_x) that would otherwise deactivate the chlorine^{3,4}. Although the heterogeneous chemistry is now well established, the composition of the clouds themselves is uncertain. It is commonly thought that they are composed of nitric acid trihydrate³, although observations have left this question unresolved^{5–14}. Here we reanalyse infrared spectra of type I PSCs obtained in Antarctica in September 1987^{15,16}, using recently measured optical constants of the various compounds that might be present in PSCs¹⁷. We find that these PSCs were not composed of nitric acid trihydrate but instead had a more complex composition, perhaps that of a ternary solution. Because cloud formation is sensitive to their composition, this finding will alter our under-

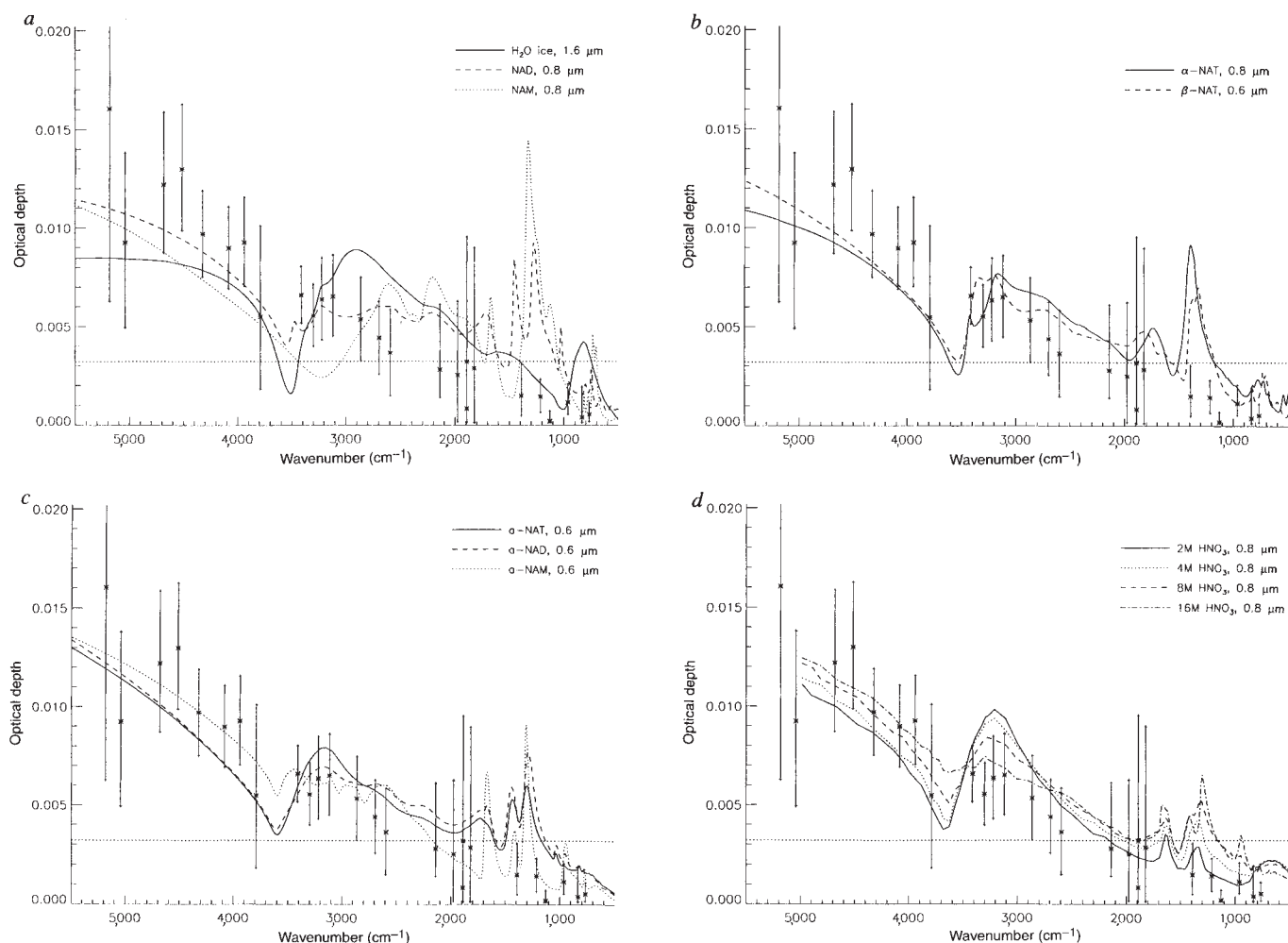


FIG. 1 *a*, Computed spectra for ice, NAM and NAD are compared with an observed spectrum (data points, with error bars) of PSCs. The numbers in the key correspond to the radius of the log-normal distribution used in the calculated spectrum. The radius shown is that which gave the best fit to the observations for that composition. The dotted horizontal line is an estimate of the instrument noise level¹⁵. Data below that level may represent random errors in the measurements, rather than

a true value of the infrared optical depth. The error bars are 1σ based on all of the frequencies included within a spectral window. *b*, As for *a* except that α -NAT and β -NAT are considered. *c*, As for *a* except that α -NAT, α -NAD and α -NAM are considered. *d*, As for *a* except that computed spectra using room-temperature optical constants²⁶ for 2 M (12 wt%), 4 M (22 wt%), 8 M (40 wt%) and 15.7 M (70 wt%) liquid nitric acid solutions are compared with the observed spectrum.

standing of the locations and conditions in which PSCs form. In addition, the extent of ozone loss depends on the ability of the PSCs to remove NO_x permanently through sedimentation. The sedimentation rates depend on PSC particle size which in turn is controlled by the composition and formation mechanism¹⁴.

The type I PSC composition was initially suggested to be nitric acid trihydrate³, solid solutions of nitric acid^{3,18}, or nitric acid monohydrate¹⁹. Subsequently, nitrogen was observed as a major component of type I PSCs^{20,21}. Vapour pressure measurements²² showed that nitric acid trihydrate is the stable crystalline form of nitric acid at winter-time stratospheric temperatures and gas-phase nitric acid concentrations. However, observed formation temperatures of type I PSCs are often several degrees too low to be consistent with nitric acid trihydrate^{5,6}. Although this inconsistency could be due to nucleation barriers⁵, observations^{7,8} suggest that other forms of nitric acid including nitric acid dihydrate⁹, supercooled solutions of nitric acid^{3,5,10} and ternary solutions of sulphuric acid, nitric acid and water^{12–14} could be present. Our current spectroscopic analysis of type I PSCs over Antarctica now provides new insights into their composition.

Infrared spectra of type I PSCs have been obtained over three days in September 1987^{15,16} over narrow spectral windows, free

from gaseous absorption, using a Michelson interferometer. Within each spectral window 7–500 frequencies were sampled. The typical standard deviation of clear spectra in the windows was used to establish an instrument noise level¹⁵. By comparing spectra from clear and cloudy regions, a mean cloud optical depth was determined in each spectral window. Calculations demonstrated that these PSCs could not be made of water ice¹⁵. However, lack of available optical constants at that time prevented a determination of whether the particles were made of crystalline forms of nitric acid, which are generally the favoured candidates for PSC composition^{3,5,7,9,22}. Here we reanalyse these spectra using newly measured optical constants of water ice, nitric acid trihydrate (in the α -NAT and β -NAT forms), nitric acid dihydrate (NAD), and nitric acid monohydrate (NAM)¹⁷. We also investigate amorphous solid solutions, α -NAT, α -NAD and α -NAM, with compositions that crystallize upon heating to form hydrates.

Because the various observed spectra are similar, we reanalyse an average of four spectra for 21 September 1987 (Fig. 1). Two spectra were obtained at 14:08–14:10 universal time (UT) near 105° W longitude and 84° S latitude, and two spectra were obtained at 14:33–14:36 UT at 90° W and 87° S. During the optical-depth analysis, these spectra were normalized to clear-

sky spectra taken at nearby locations¹⁵. We chose these four spectra because their signal-to-noise ratio was high relative to other spectra. The high signals were due to the relatively large cloud optical depths, and because the Sun was close to the horizon during the measurements which increased the optical path length through the cloud. Lidar data suggest these PSCs were at 15–18 km altitude¹⁵. Analyses of the observed spectral dependence of the infrared optical depths suggests that the radius of these particles was $\sim 0.8 \mu\text{m}$, their number concentration was $\sim 1 \text{ cm}^{-3}$, and their mass mixing ratio was $\sim 10 \text{ p.p.b.}$ (ref. 15). None of these derived particle properties is very sensitive to particle composition, as they are based on a portion of the spectrum in which scattering dominates, as discussed further below. Observations of the mass of sulphuric acid particles above Antarctica in 1987 suggest that $\sim 5\%$ of the mass of the PSCs could be sulphuric acid²³. There were ~ 8 sulphuric acid particles per cm^3 in the altitude range of the PSCs observed²⁴. The smaller number of PSC particles observed may represent lack of sensitivity of the infrared observations to small particles, rather than incomplete nucleation of the sulphuric acid particles to form PSC particles. As the sulphuric acid particles have low mass they do not have an observable effect on the infrared spectra, and because they should be uniformly distributed their spectral features would be removed when our PSC spectra are normalized to clear-sky spectra to obtain the PSC optical depths¹⁵.

The observed spectrum at frequencies above $3,500 \text{ cm}^{-1}$ (Fig. 1) is dominated by scattering of the solar beam, because micrometre-sized ice and nitric acid particles are nearly transparent at these frequencies. The slope of the spectrum at these frequencies is controlled by the particle size. Below $3,500 \text{ cm}^{-1}$ the spectrum is sensitive to particle size, but also to the optical constants of the materials because ice and nitric acid particles are absorbing at these frequencies. Varying the number of cloud particles results in a wavelength-independent scaling factor for the optical depths. We performed Mie scattering calculations of the optical depth for each set of optical constants, assuming a log-normal size distribution for the particles with a standard deviation $\sigma = 1.5$ and variable mode size. The optical depths for a given mode size were then scaled to the data, and the mode size varied to obtain the least mean-square deviation between the calculated and observed optical depths.

Figure 1a shows that the observed PSCs are not composed of ice. Ice has strong absorption features near $3,000$ and 800 cm^{-1} , which are absent in the data. We have minimized the strength of these features in the calculated optical depth spectra by making the ice particles relatively large. Increasing the particle size to produce a still better fit at $3,000 \text{ cm}^{-1}$ produces a poorer fit¹⁵ at frequencies above $4,000 \text{ cm}^{-1}$. NAM has absorption features between $3,000$ and $2,000 \text{ cm}^{-1}$, and near $1,200 \text{ cm}^{-1}$ that do not appear in the data, and NAM is not absorbing enough near $3,300 \text{ cm}^{-1}$ to fit the data (Fig. 1a). NAD does not fit the spectrum well between $3,500$ and $2,000 \text{ cm}^{-1}$, and has bands near $1,400 \text{ cm}^{-1}$ and $1,200 \text{ cm}^{-1}$ that are not observed (Fig. 1a). Figure 1b shows that α -NAT and β -NAT have strong bands at $1,200$ and $1,400 \text{ cm}^{-1}$ which are not observed. Figure 1c shows that α -NAT and α -NAD also have strong bands near $1,400$ and $1,200 \text{ cm}^{-1}$ that are not observed. In general, all of the constituents whose spectra have been calculated have strong absorption bands that are not observed. In particular, NAT has bands that are calculated to lie more than two standard deviations above the observed spectra.

We performed many additional comparisons between observed and calculated spectra. We varied the particle size and the width of the size distribution, tried mixtures in which different composition particles either coexisted or were in cloud layers at different altitudes, explored shell and core combinations of materials, and computed spectra for particles with oblate or prolate shapes. None of these attempts resulted in a better match between calculated and observed spectra than those shown in Fig. 1.

The calculated spectrum of α -NAM with particle sizes of $0.6 \pm 0.2 \mu\text{m}$ does agree with the observations (Fig. 1c) because α -NAM has relatively low imaginary refractive indices between its major absorption bands from $1,000$ to $1,500 \text{ cm}^{-1}$. The major objection to α -NAM as a constituent of PSCs is that its nitric acid vapour pressure is too large for it to occur in the stratosphere¹⁰. Moreover, we have examined the observed infrared extinction in the region near $1,300 \text{ cm}^{-1}$, and do not find evidence for the strong band which α -NAM has at that frequency.

Recent laboratory and theoretical work points to the possible existence of ternary solutions of nitric acid, sulphuric acid and water in PSCs^{12, 14}. Given the low temperature at the locale of our observations (near or below 192 K), as well as the dehydrated and denitrified conditions of the lower stratosphere in September 1987, theory suggests that supercooled ternary solutions would be $\sim 5 \text{ wt\%}$ sulphuric acid and $30\text{--}50 \text{ wt\%}$ nitric acid¹⁴. The refractive indices for amorphous nitric acid compounds discussed in this Letter, which should be similar to those of supercooled nitric acid²⁵, are available only for solutions more concentrated than $50 \text{ wt\%}$. But room-temperature refractive indices for less-concentrated nitric acid solutions are available²⁶, and Fig. 1d shows that these provide a much better fit to the observations than any of the materials investigated in Fig. 1a–c. As the nitric acid concentration declines, as shown in Fig. 1d, the nitrate bands near $1,200\text{--}1,400 \text{ cm}^{-1}$ decrease in strength. A solution less concentrated than 4 M ($22 \text{ wt\%}$) is required to fit the spectrum. However, the $3,000 \text{ cm}^{-1}$ water band increases in strength as the nitric acid concentration declines. A solution more concentrated than 4 M is required to fit the spectrum above $3,000 \text{ cm}^{-1}$. At present no optical constants are available either for ternary solutions, or for binary nitric acid solutions with relevant compositions at stratospheric temperatures. Studies of sulphuric acid binary solutions (which may have similarities to nitric acid solutions), have found that the infrared spectra of solutions containing $<50\%$ sulphuric acid are strongly temperature dependent²⁷. Quantitative determination of the nitric acid concentration from the observed atmospheric spectra will require measurements of the low-temperature optical constants of ternary mixtures. Additional atmospheric spectra would also be useful to search for changes in the composition of the particles caused by variations in environmental conditions¹⁴.

It has been suggested that there is a hemispheric asymmetry in PSC composition, with NAT clouds existing in the Antarctic, but other compounds in the Arctic⁵. However, our results show that the type I PSCs over Antarctica during September 1987, a period with significant ozone loss, were not composed of NAT. Our results imply that it is much more difficult to form NAT than thought previously. Of course, we cannot rule out NAT formation at other times and places than the ones we studied.

The extent of future ozone loss is controlled in part by the ability of PSCs to permanently remove NO_x from the stratosphere, so-called 'denitrification'. Large particles with rapid sedimentation rates form when a small fraction of the pre-existing particles nucleate to form crystals such as NAT. In contrast, small particles form when liquid PSCs grow within all of the stratospheric aerosols¹⁴. Most theories for future ozone loss assume that NAT is the principal component of PSCs. If so, stratospheric cooling during the next century due to the greenhouse effect may trigger an increase in the formation of NAT clouds in the colder Arctic stratosphere. The resulting enhanced sedimentation of nitric acid has been predicted to reduce the sensitivity of stratospheric ozone to anthropogenic perturbations in nitrogen oxides²⁸. But if NAT is as difficult to form as our study implies, then only smaller ternary solution particles may form as the stratosphere cools. These smaller particles would not efficiently remove nitric acid from the stratosphere at temperatures above the frost point.

Massive chlorine-catalysed ozone loss such as observed in the Antarctic ozone hole may require denitrification. Denitrification

of the stratosphere by ice clouds (type II PSCs) is inefficient if type I PSCs contain NAT²⁹. However, the higher vapour pressures of ternary solutions would greatly speed the transfer of nitric acid from type I to type II PSCs. Therefore the lack of evidence for type I PSCs containing NAT implies that nitrogen removal from the stratosphere may be more closely related to the formation of ice particles than it is to nitric acid particles³.

It has also been suggested that enhancements in the concentrations of stratospheric nitric acid and water vapour from future fleets of supersonic aircraft might lead to significant expansions of type I PSCs composed of NAT³⁰. But the variation in PSC abundance for a given change in nitric acid vapour supply may be quite different for NAT than for other PSC components. NAT formation should occur at a definite onset temperature, whereas particles composed of ternary solutions increase gradually in size and coverage as the concentration of nitric acid vapour increases. Assuming a NAT composition, regions currently near the NAT formation point would experience a sudden drastic increase in PSC abundance if the nitric acid vapour concentration was increased past the threshold value. In contrast, only a modest change in the abundance of ternary aerosols would occur as the existing aerosols would swell as conditions changed. In areas far from the NAT threshold, small changes in the nitric acid vapour abundance would have no effect on NAT aerosols, but the ternary aerosols would change their size and hence their potential to cause ozone loss.

Owing to the complexity of stratospheric chemistry and PSC microphysics, it is not possible to anticipate the quantitative changes that might occur in theoretical studies which have assumed PSCs are composed of NAT. But it is clear that the properties of other possible PSC constituents are sufficiently

different from those of NAT that future theories should not treat PSCs as though they were composed only of NAT. It is also clear that we need to understand better the environmental conditions that control PSC composition. □

Received 15 September 1994; accepted 29 March 1995.

1. Solomon, S. *Nature* **347**, 347–354 (1990).
2. Toon, O. B. & Turco, R. P. *Scient. Amer.* **264**, 68–74 (1991).
3. Toon, O. B., Hamill, P., Turco, R. P. & Pinto, J. *Geophys. Res. Lett.* **13**, 1284–1287 (1986).
4. Schoeberl, M. R. et al. *Geophys. Res. Lett.* **20**, 2511–2514 (1993).
5. Dye, J. E. et al. *J. geophys. Res.* **97**, 8015–8034 (1992).
6. Kawa, S. R. et al. *J. geophys. Res.* **97**, 7925–7938 (1992).
7. Browell, E. V. et al. *Geophys. Res. Lett.* **17**, 385–388 (1990).
8. Rosen, J. M., Oltmans, S. J. & Evans, W. F. *Geophys. Res. Lett.* **16**, 791–794 (1989).
9. Worsnop, D. R., Fox, L. E., Zahniser, M. S. & Wofsy, S. C. *Science* **259**, 71–74 (1993).
10. Hanson, D. R. *Geophys. Res. Lett.* **17**, 421–424 (1990).
11. Toon, O. B. et al. *Science* **261**, 1136–1140 (1993).
12. Molina, M. J. et al. *Science* **261**, 1418–1423 (1993).
13. Carslaw, K. S. et al. *Geophys. Res. Lett.* **21**, 2479–2482 (1994).
14. Tabazadeh, A. et al. *Geophys. Res. Lett.* **21**, 1619–1622 (1994).
15. Kinne, S. et al. *J. geophys. Res.* **94**, 16481–16491 (1989).
16. Toon, G. C. et al. *J. geophys. Res.* **94**, 16571–16596 (1989).
17. Toon, O. B., Tolbert, M. A., Koehler, B. G., Middlebrook, A. M. & Jordan, J. J. *Geophys. Res.* **99**, 25631–25654 (1994).
18. Crutzen, P. J. & Arnold, F. *Nature* **324**, 651–655 (1986).
19. McElroy, M. B., Salawitch, R. J. & Wofsy, S. C. *Geophys. Res. Lett.* **13**, 1296–1299 (1986).
20. Fahey, D. W. et al. *J. geophys. Res.* **94**, 11299–11315 (1989).
21. Pueschel, R. F. et al. *J. geophys. Res.* **94**, 11271–11284 (1989).
22. Hanson, D. & Mauersberger, K. *Geophys. Res. Lett.* **15**, 855–858 (1988).
23. Gandrud, B. W. et al. *J. geophys. Res.* **94**, 11285–11297 (1989).
24. Hofmann, D. J., Rosen, J. M., Harder, J. W. & Hereford, J. V. *J. geophys. Res.* **94**, 11253–11270 (1989).
25. Koehler, B. G., Middlebrook, A. M. & Tolbert, M. A. *J. geophys. Res.* **97**, 8065–8074 (1992).
26. Querry, M. R. & Tyler, I. L. *J. chem. Phys.* **72**, 2495–2499 (1980).
27. Zhang, R., Wooldridge, P. J., Abbatt, J. P. D. & Molina, M. J. *J. phys. Chem.* **97**, 7351–7358 (1993).
28. Pitari, G. & Ricciardulli, L. *Geophys. Res. Lett.* **21**, 1791–1794 (1994).
29. Toon, O. B., Turco, R. P. & Hamill, P. *Geophys. Res. Lett.* **17**, 445–448 (1990).
30. Peter, Th., Bruhl, C. & Crutzen, P. J. *Geophys. Res. Lett.* **18**, 1465–1468 (1991).

ACKNOWLEDGEMENTS. We thank S. Kinne and G. Toon for providing the infrared spectra, A. Tabazadeh, E. Jensen and A. Middlebrook for helpful comments on the work and J. Jordan for assisting with our analysis. This work was supported by NASA and the US NSF.

Influence of static stress changes on earthquake locations in southern California

Ruth A. Harris, Robert W. Simpson & Paul A. Reasenber

US Geological Survey, Mail Stop 977, 345 Middlefield Road, Menlo Park, California 94025, USA

EARTHQUAKES induce changes in static stress on neighbouring faults that may delay, hasten or even trigger subsequent earthquakes^{1–10}. The length of time over which such effects persist has a bearing on the potential contribution of stress analyses to earthquake hazard assessment, but is presently unknown. Here we use an elastic half-space model¹¹ to estimate the static stress changes generated by damaging (magnitude $M \geq 5$) earthquakes in southern California over the past 26 years, and to investigate the influence of these changes on subsequent earthquake activity. We find that, in the 1.5-year period following a $M \geq 5$ earthquake, any subsequent nearby $M \geq 5$ earthquake almost always ruptures a fault that is loaded towards failure by the first earthquake. After this period, damaging earthquakes are equally likely to rupture loaded and relaxed faults. Our results suggest that there is a short period of time following a damaging earthquake in southern California in which simple Coulomb failure stress models could be used to identify regions of increased seismic hazard.

Do static stress changes trigger earthquakes, and, if so, over what distances and times can triggering occur? We attempt to answer these questions by examining the $M \geq 5$ earthquakes (55

in number) that have occurred in southern California since 1968 (Fig. 1). Based on studies of aftershock sequences in California, the probability that a $M \geq 5$ California earthquake will follow a $M = 6$ earthquake within 3 years is 0.77 (refs 12, 25), but this does not offer guidance as to the locations of future earthquakes, information that would be useful for mitigating seismic hazard.

We tested the hypothesis that static stress changes generated by $M \geq 5$ southern California earthquakes influence subsequent $M \geq 5$ events, by evaluating the stress interactions between the earthquakes. We used Okada's equations for dislocations in an elastic half-space¹¹ to calculate the static stress changes produced by a rectangular dislocation (slip) model of each earthquake. The stress changes were calculated at each subsequent earthquake's hypocentre and resolved onto the subsequent earthquake's fault plane. The change in normal stress, $\Delta\sigma_n$ (tension positive), and the change in shear stress resolved in the slip direction of the subsequent earthquake, $\Delta\tau_s$, were used to determine the change in Coulomb failure stress, ΔCFS , given by

$$\Delta CFS = \Delta\tau_s + \mu' \Delta\sigma_n \quad (1)$$

where μ' is the apparent coefficient of friction, which includes the effects of pore pressure^{6,9}. Assuming that the Coulomb failure criterion applies, earthquakes that receive positive ΔCFS at their hypocentres would be more loaded towards failure and those receiving negative ΔCFS more relaxed.

We initially studied 44 $M \geq 5$ southern California earthquakes that occurred between 1968 and 1993 (Table 1), and evaluated the stress interactions between the 946 unique pairs of events. (The 1971 San Fernando earthquake's four $M \geq 5$ aftershocks, all within 5 km of the mainshock, were not included in our study because not enough is known about them to calculate ΔCFS .) We eliminated from consideration 907 pairs for which $|\Delta CFS| < 0.1$ bar. (0.1 bar exceeds tidal stress levels by a factor

TABLE 1 44 $M \geq 5$ earthquakes 1968–93

Date (yr mth d)	Location (lat. N, long. W)	Mag.	Name
68 04 09	33° 11.40', 116° 07.72'	6.5	Borrego Mountain
68 07 05	34° 07.06', 119° 42.15'	5.3	Santa Barbara Channel
69 01 23	33° 53.21', 116° 02.42'	5.0	
69 04 28	33° 20.60', 116° 20.78'	5.8	Coyote Mountain
70 09 12	34° 16.19', 117° 32.40'	5.2	Lytle Creek
71 02 09	34° 24.67', 118° 24.04'	6.6	San Fernando
71 09 30	33° 02.01', 115° 49.24'	5.0	
73 02 21	34° 03.89', 119° 02.10'	5.5	Point Magu
73 08 06	33° 59.16', 119° 28.52'	5.0	
75 06 01	34° 30.94', 116° 29.73'	5.0	Galway Lakes
76 11 04	33° 07.89', 115° 37.40'	5.0	
78 08 13	34° 20.82', 119° 41.76'	5.1	Santa Barbara
79 01 01	33° 56.66', 118° 40.88'	5.2	Malibu
79 03 15	34° 19.64', 116° 26.69'	5.3	Homestead Valley
79 10 15	32° 36.82', 115° 19.09'	6.4	Imperial Valley
80 02 25	33° 30.06', 116° 30.75'	5.5	Anza
81 04 26	33° 05.90', 115° 37.90'	5.7	Westmorland
81 09 04	33° 40.26', 119° 06.67'	5.5	Santa Barbara Island
86 07 08	33° 59.95', 116° 36.51'	5.6	North Palm Springs
86 07 13	32° 57.96', 117° 52.32'	5.4	Oceanside
87 10 01	34° 03.68', 118° 04.71'	5.9	Whittier Narrows
87 10 04	34° 04.42', 118° 05.88'	5.3	
87 11 24	33° 04.95', 115° 46.51'	6.2	Elmore Ranch
87 11 24	33° 00.69', 115° 51.31'	6.6	Superstition Hills
88 06 10	34° 56.58', 118° 44.56'	5.4	
88 12 03	34° 08.93', 118° 08.08'	5.0	Pasadena
89 01 19	33° 55.12', 118° 37.64'	5.0	
90 02 28	34° 08.29', 117° 42.17'	5.3	Upland
91 06 28	34° 15.69', 117° 59.97'	5.8	Sierra Madre
92 04 23	33° 57.33', 116° 17.97'	6.1	Joshua Tree
92 06 28	34° 12.13', 116° 25.95'	7.3	Landers
92 06 28	34° 09.96', 116° 51.34'	5.3	JT-L seq.
92 06 28	34° 09.94', 116° 49.35'	6.2	Big Bear
92 06 29	34° 06.21', 116° 23.99'	5.4	JT-L seq.
92 06 29	33° 52.11', 116° 16.09'	5.2	JT-L seq.
92 07 01	34° 20.01', 116° 27.59'	5.4	JT-L seq.
92 07 05	34° 34.81', 116° 19.10'	5.4	Pisgah
92 07 11	35° 12.60', 118° 03.94'	5.7	Mojave
92 07 24	33° 53.88', 116° 16.88'	5.0	JT-L seq.
92 08 17	34° 11.71', 116° 51.63'	5.2	JT-L seq.
92 09 15	34° 03.64', 116° 21.50'	5.2	JT-L seq.
92 11 27	34° 20.54', 116° 54.20'	5.3	JT-L seq.
92 12 04	34° 21.98', 116° 54.19'	5.1	JT-L seq.
93 05 28	35° 08.96', 119° 06.22'	5.2	White Wolf

of 3–10.) This threshold is based on a study⁴ of the effects of the 1989 $M=7.1$ Loma Prieta earthquake in central California, where a significant correlation between microseismicity rate changes and static stress changes held at the $|\Delta\text{CFS}| \geq 0.1$ bar level. We also eliminated 12 earthquake pairs closer than 5 km hypocentral distance because the ΔCFS calculations would have required unknown fine details of the slip distributions for each earthquake. Our distance and ΔCFS selection criteria provided 27 pairs of $M \geq 5$ southern California earthquakes for which static stress calculations seemed meaningful. For these 27 pairs the smallest first earthquake in the pair had $M=5.4$. The greatest hypocentral distance for a pair was 40 km.

The temporal pattern of the 27 pairs is shown in Fig. 2a. 16 pairs are characterized by a time delay of less than or equal to 1.5 years, of which 15 have positive ΔCFS (Table 2). Twelve of these pairs involved earthquakes in the 1992 Joshua Tree–Landers earthquake sequence¹³, although some intuitive pairs, such as Joshua Tree–Landers mainshock are not in our list because $\Delta\text{CFS} < 0.1$ bar. After the first 1.5 years, there are approximately equal numbers of positive and negative ΔCFS pairs, although the numbers are small. Thus, any long-term (>1.5 years) influence of static stress changes on future earthquakes is not apparent in our sample. We have also explored the effects of varying μ' and found that similar results hold for all values of this parameter, although there are more positive pairs for low values of μ' (Fig. 2b).

Could this correlation between positive ΔCFS and subsequent earthquakes be explained by a systematic bias? For example, there could be a geometrical bias for positive ΔCFS pairs: two earthquakes on a single planar fault yield $\Delta\text{CFS} > 0$ regardless of whether or not static stress changes from the first earthquake triggered the second. We tested the possibility that nearby faults tend to be oriented to produce a positive ΔCFS pair by calculating ΔCFS on all known active Quaternary faults¹⁴ near three $M > 5$ earthquakes (1968 Borrego Mountain, 1992 Landers and 1994 Northridge). We found that 50% or more of the fault segments near each of these earthquakes had negative ΔCFS . If $M > 5$ earthquake pairs were occurring randomly irrespective of static stress changes, half would be expected to occur on faults that were loaded, and half on faults that were relaxed. This is not observed in our data. Based on our small sample, almost all pairs close enough to be plausibly triggered by static stress

FIG. 1 Map of southern California, showing locations of $M \geq 5$ earthquakes between the years 1968 and 1994^{13,21} (L. Jones, K. Hutton and E. Hauksson, personal communication). NR points to the 17 January 1994 Northridge mainshock. The beach balls show the focal mechanisms of the earthquakes. Larger beach balls represent higher-magnitude earthquakes. Open circles represent earthquakes with unknown focal mechanisms.

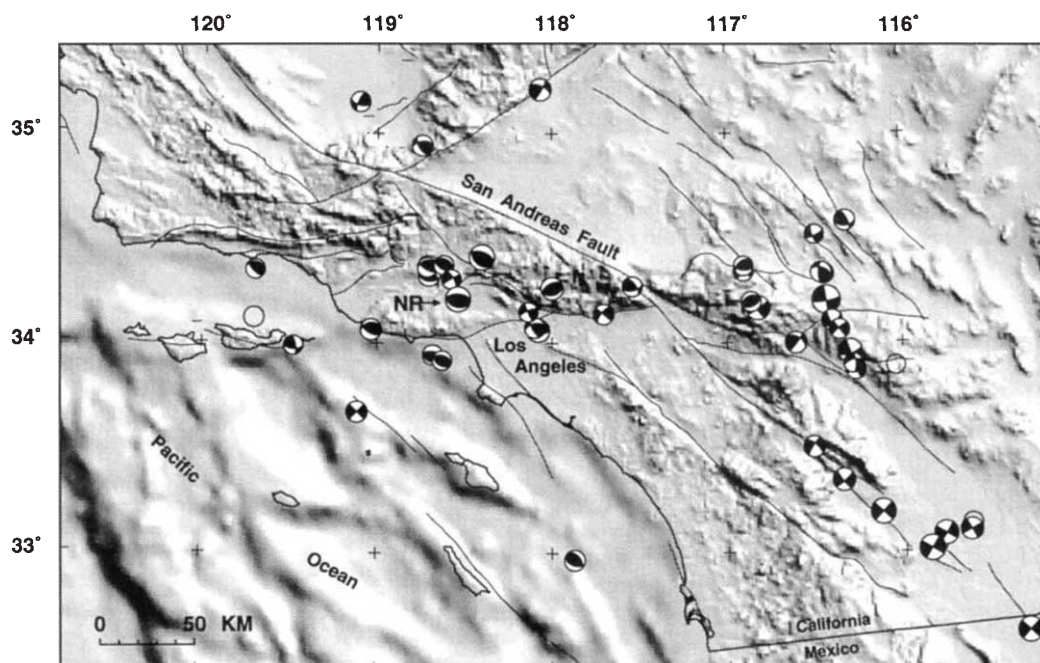
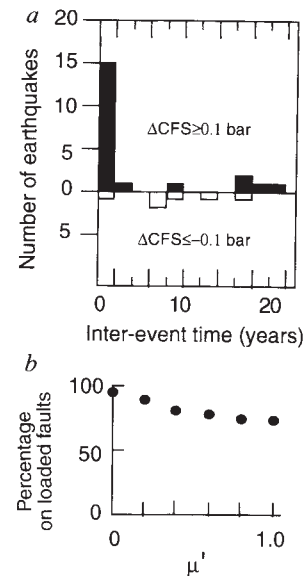


FIG. 2 *a*, Distribution of ΔCFS versus time between the earthquakes for the 27 selected pairs of earthquakes from 1968 to 1993. Assuming $\mu' = 0.0$ (ref. 22), within the first 1.5-year interval, 15 earthquakes occurred on fault planes that were loaded ($\Delta CFS \geq 0.1$ bar) by a previous earthquake, and only one occurred on a fault plane that was relaxed ($\Delta CFS \leq -0.1$ bar) by a previous event. These results change to 14 earthquakes on loaded faults and 2 earthquakes on relaxed faults if $\mu' = 0.2$ (see *b*). After 1.5 years, $M \geq 5$ earthquakes are about equally likely to rupture a fault plane that was loaded or relaxed by a preceding earthquake. This is demonstrated by the approximately equal number of loaded ($\Delta CFS \geq 0.1$ bar) and relaxed ($\Delta CFS \leq -0.1$ bar) earthquakes for >1.5 years inter-event time. *b*, Effect of μ' on our results. The percentage of $M \geq 5$ earthquakes rupturing fault planes loaded ($\Delta CFS \geq 0.1$ bar) by an earthquake within the previous 1.5 years is plotted against μ' . The positive correlation between earthquakes and loaded faults also holds for high values of the coefficient of friction, although the match is better for low values of μ' . A low value for μ' is consistent with analyses of static stress changes and aftershocks of the 1992 Landers earthquake^{23,24} and with earthquakes in central California⁴.

TABLE 2 16 earthquake pairs 1968–93 with time ≤ 1.5 years

Date (yr mth d)	Location (lat. N, long. W)	Depth (km)	Mag.	Time (yr)	Dist. (km)	ΔCFS (bar)	Name*
68 04 09	33° 11.40', 116° 07.72'	11.0	6.5				Borrego Mtn
69 04 28	33° 20.60', 116° 20.78'	20.0	5.8	1.0	26	+0.3	Coyote Mtn
79 10 15	32° 36.82', 115° 19.09'	12.1	6.4				Imperial Vly
81 04 26	33° 05.90', 115° 37.90'	3.0	5.7	1.5	20	+0.4	Westmorland
87 10 01	34° 03.68', 118° 04.71'	14.7	5.9				Whit. Narrows
88 12 03	34° 08.93', 118° 08.08'	15.5	5.0	1.2	11	+0.3	Pasadena
87 11 24	33° 04.95', 115° 46.51'	10.6	6.2				Elmore Ranch
87 11 24	33° 00.69', 115° 51.31'	1.7	6.6	0.0	11	+1.0	Sup. Hills
92 04 23	33° 57.33', 116° 17.97'	14.4	6.1				Joshua Tree
92 06 29	34° 06.21', 116° 23.99'	10.5	5.4	0.2	19	+0.2	JT–L seq.
92 04 23	33° 57.33', 116° 17.97'	14.4	6.1				Joshua Tree
92 06 29	33° 52.11', 116° 16.09'	0.5	5.2	0.2	10	+0.5	JT–L seq.
92 04 23	33° 57.33', 116° 17.97'	14.4	6.1				Joshua Tree
92 07 24	33° 53.88', 116° 16.88'	9.6	5.0	0.3	7	+2.2	JT–L seq.
92 04 23	33° 57.33', 116° 17.97'	14.4	6.1				Joshua Tree
92 09 15	34° 03.64', 116° 21.50'	9.5	5.2	0.4	13	+0.6	JT–L seq.
92 06 28	34° 12.13', 116° 25.95'	3.2	7.3				Landers
92 06 28	34° 09.94', 116° 49.35'	12.9	6.2	0.0	37	+0.7	Big Bear
92 06 28	34° 12.13', 116° 25.95'	3.2	7.3				Landers
92 06 29	33° 52.11', 116° 16.09'	0.5	5.2	0.0	20	+1.2	JT–L seq.
92 06 28	34° 12.13', 116° 25.95'	3.2	7.3				Landers
92 07 05	34° 34.81', 116° 19.10'	7.5	5.4	0.0	21	–2.2	Pisgah
92 06 28	34° 12.13', 116° 25.95'	3.2	7.3				Landers
92 07 24	33° 53.88', 116° 16.88'	9.6	5.0	0.1	17	+1.2	JT–L seq.
92 06 28	34° 12.13', 116° 25.95'	3.2	7.3				Landers
92 08 17	34° 11.71', 116° 51.63'	12.6	5.2	0.1	40	+0.2	JT–L seq.
92 06 28	34° 12.13', 116° 25.95'	3.2	7.3				Landers
92 12 04	34° 21.98', 116° 54.19'	1.8	5.1	0.4	37	+0.3	JT–L seq.
92 06 28	34° 09.94', 116° 49.35'	12.9	6.2				Big Bear
92 11 27	34° 20.54', 116° 54.20'	1.4	5.3	0.4	21	+0.1	JT–L seq.
92 06 29	34° 06.21', 116° 23.99'	10.5	5.4				JT–L seq.
92 09 15	34° 03.64', 116° 21.50'	9.5	5.2	0.2	6	+0.7	JT–L seq.

* Abbreviations used; Whit., Whittier; Sup., Superstition; JT–L seq., Joshua Tree–Landers sequence.

changes ($|\Delta\text{CFS}| \geq 0.1$ bar) have positive values of ΔCFS , and almost no pairs have negative values of ΔCFS . This correlation of positive ΔCFS and earthquakes suggests, regardless of physical cause, that faults within 40 km of an earthquake that are oriented to produce negative ΔCFS are less likely to be at risk than faults oriented to produce positive ΔCFS .

The 17 January 1994 Northridge, California, $M=6.7$ earthquake¹⁵ provided additional data. We represented the Northridge mainshock by an 8-km-long, 10-km-tall, south-dipping rectangular fault plane with 3.5 m of reverse slip¹⁶. We used this model of the mainshock to calculate the resulting tractions at the hypocentres of the six $M \geq 5$ aftershocks (E. Hauksson, personal communication). All of the aftershocks were at least 5 km from the rectangular dislocation surface representing the first event. We calculated a ΔCFS for each aftershock by resolving the induced stress changes in the aftershock's slip direction. We assumed low values for μ' of 0.0–0.2 (ref. 4). All six of the $M \geq 5$ Northridge aftershocks ruptured fault planes that were loaded by >0.1 bar by the Northridge mainshock. The Northridge $M \geq 5$ data, which are not plotted in Fig. 2a, enhance our conclusion.

We have provided evidence that the second earthquake in a pair (distance >5 km, $|\Delta\text{CFS}| \geq 0.1$ bar) of closely timed (≤ 1.5 years) $M \geq 5$ southern California earthquakes almost always ruptures a fault plane that experienced an increase in Coulomb failure stress. This conclusion reinforces previous static stress studies that correlate patterns of increased seismicity with regions of increased Coulomb failure stress. Our study differs from most efforts in that we have calculated ΔCFS on the fault planes and in the slip directions of the earthquakes¹⁷.

Physical explanations for the time-dependence in Fig. 2a include traditional explanations for the time delay of aftershocks: fluid diffusion, aseismic slip at depth, viscoelastic relaxation and static fatigue¹⁸. Dieterich has recently derived¹⁹ an expression for duration time for aftershocks, defined as the time for the rate of aftershocks to return to background seismicity rates, which is based on a laboratory-determined rate and state-dependent friction law. If we use this formulation to evaluate the case of southern California earthquake pairs, where the second earthquake is a $M=5$ event, we obtain pair duration times of the order of 1–2 years. This seems to be consistent with our results for southern California.

Analysis of 26 years of $M \geq 5$ earthquakes in southern California suggests that obvious evidence for $M \geq 5$ earthquake-triggering, consistent with static stress increases, occurs only in the first 1.5 years after an event. We cannot rule out the possibility that second earthquakes, in pairs with positive ΔCFS and time delays of >1.5 years, were also advanced by the added stress load²⁰, but this is not apparent in our sample of southern California earthquakes. After 1.5 years, the correlation between $M \geq 5$ earthquakes and static stress increases induced by previous events is probably masked by the complex tectonic processes that generate earthquakes at an active plate boundary. $\square$

Received 30 December 1994; accepted 29 March 1995.

- Das, S. & Scholz, C. *Bull. seism. Soc. Am.* **71**, 1669–1675 (1981).
- Stein, R. S. & Lisowski, M. *J. geophys. Res.* **88**, 6477–6490 (1983).
- Hudnut, K. W., Seeber, L. & Pacheco, J. *Geophys. Res. Lett.* **16**, 199–202 (1989).
- Reasenber, P. A. & Simpson, R. W. *Science* **255**, 1687–1690 (1992).
- Du, Y. & Aydin, A. *J. geophys. Res.* **98**, 9947–9962 (1993).
- Harris, R. A. & Simpson, R. W. *Nature* **360**, 251–254 (1992).
- Joumé, S. C. & Sykes, L. R. *Science* **258**, 1325–1328 (1992).
- Stein, R. S., King, G. C. P. & Lin, J. *Science* **258**, 1328–1332 (1992).
- Simpson, R. W. & Reasenber, P. A. *Prof. Pap.* 1550-F, 55–89 (US Geol. Surv., Washington DC, 1994).
- Harris, R. A. & Simpson, R. W. (abstr.) *Eos* **74**, 427 (1993).
- Okada, Y. *Bull. seism. Soc. Am.* **82**, 1018–1040 (1992).
- Reasenber, P. A. & Jones, L. M. *Science* **243**, 1173–1176 (1989).
- Hauksson, E., Jones, L. M., Hutton, K. & Eberhart-Phillips, D. J. *Geophys. Res.* **98**, 19835–19858 (1993).
- Jennings, C. W. O.F. Rep. 92-03 (Calif. Dept of Conserv. Div. of Mines and Geol., Sacramento, 1992).
- Scientists of the US Geological Survey and Southern California Earthquake Center *Science* **266**, 389–397 (1994).

- Hudnut, K. W. et al. in *Abstr 89th Meeting of the Seism. Soc. Am.* 17 (Seism. Soc. Am., Pasadena, 1994).
- Armbruster, J. G. & Seeber, L. (abstr.) *Eos* **72**, 310 (1991).
- Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).
- Dieterich, J. J. *geophys. Res.* **99**, 2601–2618 (1994).
- Stein, R. S., King, G. C. P. & Lin, J. *Science* **265**, 1432–1435 (1994).
- Hutton, L. K. & Jones, L. M. *Bull. seism. Soc. Am.* **83**, 313–329 (1993).
- Kagan, Y. Y. *Geophys. J. Int.* **117**, 345–364 (1994).
- Gross, S. J. *Seism. Res. Lett.* **65**, 53 (1994).
- Harris, R. A. & Simpson, R. W. *Seism. Res. Lett.* **65**, 54 (1994).
- Reasenber, P. A. & Jones, L. M. *Science* **265**, 1251–1252 (1994).

ACKNOWLEDGEMENTS. We thank D. Castillo and J. Dieterich for stimulating discussions, L. Jones, E. Hauksson and K. Hutton for providing earthquake focal mechanisms and locations, and M. Murray and K. Hudnut for their preliminary slip models of the Northridge earthquake. Reviews by R. Stein and D. Hill improved the manuscript. R.A.H. was partially supported by an NSF VPW fellowship at Stanford University.

An aquatic sloth from the Pliocene of Peru

C. de Muizon & H. G. McDonald

Institut Français d'Études Andines, URA 12 CNRS, casilla 18-1217, Lima 18, Peru
Hagerman Fossil Beds National Monument, PO Box 570, Hagerman, Idaho 83332, USA

GROUND sloths (Gravirada, Xenarthra) are known from middle or late Oligocene to late Pleistocene in South America¹ and from late Miocene to late Pleistocene in North America². They are medium to gigantic in size and have terrestrial³ habits. Discovery of abundant and well preserved remains of a new sloth (*Thalassocnus natans*), in marine Pliocene deposits from Peru^{4–6} drastically expands our knowledge of the range of adaptation of the order. The abundance of individuals, the absence of other land mammals in the rich marine vertebrate fauna of the site^{5,6}, and the fact that the Peruvian coast was a desert during the Pliocene^{7,8} suggest that it was living on the shore and entered the water probably to feed upon sea-grasses or seaweeds. The morphology of premaxillae, femur, caudal vertebrae (similar to those of otters and beavers) and limb proportions are in agreement with this interpretation.

Order Xenarthra
Family Megalonychidae
Subfamily Nothrotheriinae
Thalassocnus natans, gen. et sp. nov.

Holotype. MNHN SAS 734 (Muséum national d'Histoire naturelle, Paris). A partial skeleton including skull, mandible, cervical, thoracic, lumbar and some caudal vertebrae, most ribs and some limb bones.

Locality and horizon. Sud-Sacaco^{5,6} (Department of Arequipa, Peru); Pisco Formation. Early Pliocene^{6,9,10}.

Diagnosis. Size similar to *Nothrotherium*¹¹, distinguished from *Pronothrotherium*¹² and *Nothropus*¹³ by absence of caniniform teeth: cheek dentition reduced to 4/3 as in *Nothrotherium* and *Nothrotheriops*¹⁴; premaxillae elongated in comparison to other nothrotheres, expanded at anterior margin, and deflected ventrally; nasals expanded anteriorly with continuous anterior edge; in contrast to other nothrotheres, dorsal profile of the skull is flat and ascending process of jugal extends to dorsum of skull; pronator crest of radius elongated proximally with proximal portion enlarged; pisiform mediolaterally compressed; digit three of the manus with proximal and second phalanges co-ossified as in megatheres, contrary to other nothrotheres; third trochanter forming continuous crest with greater trochanter; fovea for round ligament present as notch on posteromedial side of head; distal process of patella reduced; fifth metatarsal compressed mediolaterally; transverse processes of caudal vertebrae extending entire length of centrum, perforated by foramen and/or bifurcate in posterior caudals; tibia proportionally longer than

in other sloths, about 85% of femoral length; ventral surface of tuber calcis of calcaneum in contact with the ground in contrast to other nothotheres in which only posterior margin of tuber calcis contacts ground.

Taphonomic data suggest at least semi-aquatic habits for *T. natans*. The marine vertebrate fauna of Sud-Sacaco is extremely abundant and has yielded thousands of specimens belonging to more than 50 families including fish, crocodiles, birds and mammals^{5,6,15-17}. The only mammal belonging to an order traditionally regarded as terrestrial is *T. natans*. Articulated skeletons or partial skeletons of sloths are extremely abundant and more numerous than those of seals or dolphins, probably approaching the number of penguins. The fauna was deposited in a beach

environment¹⁸. The Peruvian coast has been a desert since the middle Miocene^{7,8}.

Considering the taphonomic arguments, several anatomical data, which clearly differ from those in other sloths, can be interpreted as aquatic adaptation. If taken alone, they would probably not be relevant because some of them could very well be regarded as fossorial adaptations. Considered within their context, they reinforce the hypothesis drawn from taphonomy.

The premaxillae of *T. natans* are long, ventrally deflected, have a spatulate transversely expanded apex, and a spongy structure indicative of a rich blood supply (Fig. 1a-c). This suggests the presence of a powerful (for a nothotherid) lip, reminiscent of sirenians. In two other specimens of *Thalassocnus*, the apex of the mandible is also broadly expanded; this feature is absent in the holotype (Fig. 1d, e). Furthermore, the attachment of premaxillae of *Thalassocnus* to maxillae, stronger than in other nothotheres¹⁹, is consistent with the strength of the lip. The expanded apex and the stoutness of the premaxillae of *Thalassocnus*, (as compared to the slender and pointed premaxillae of *Nothotherium*¹⁹ and *Nothotheriopsis*¹⁴) could be related to grazing²⁰ on sea-grasses. Other nothotheres are interpreted as browsers^{21,22}.

The femur of *Thalassocnus* (Fig. 2a) is narrow compared to other nothotheres^{14,22} and the patella^{15,23} (Fig. 2c) has a reduced

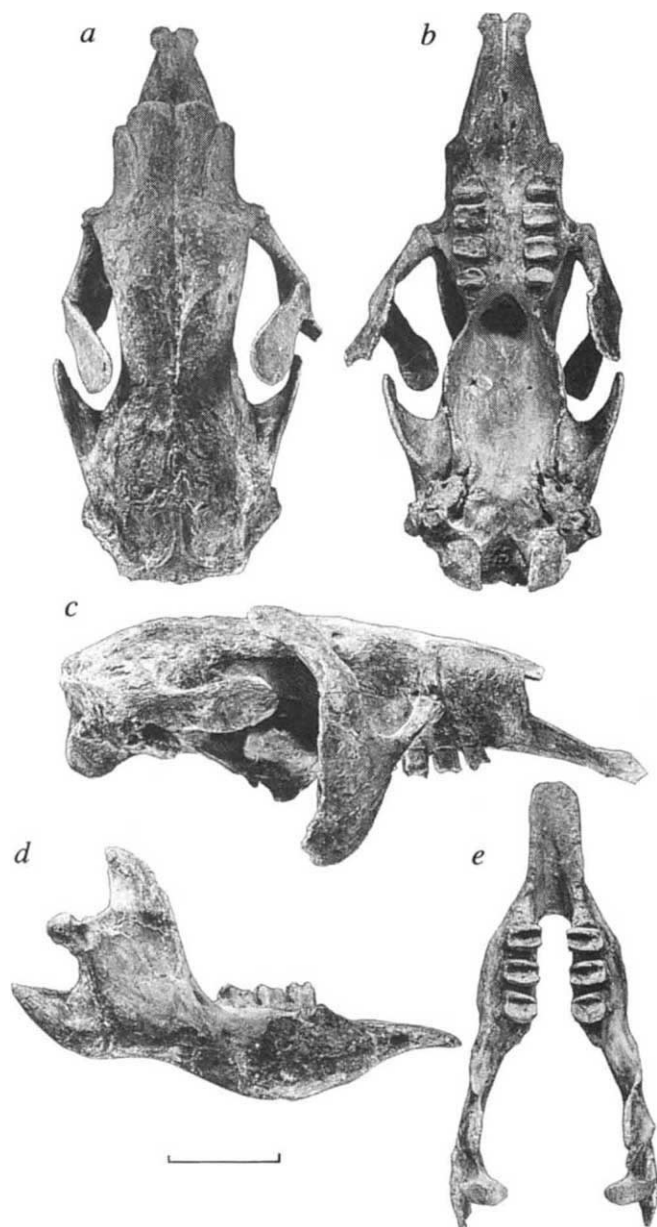


FIG. 1 *Thalassocnus natans* gen. et sp. nov. (Holotype, MNHN SAS 734). a, Skull in dorsal view; b, skull in ventral view; c, skull in lateral view; d, mandible in lateral view; e, mandible in dorsal view. In MNHN SAS 734 the apex of the mandibular symphysis is not expanded. However, an anterior expansion of the apex of the mandible is observed in a specimen from the Staatliches Museum für Naturkunde, Karlsruhe, Germany, (SMNK 1267 PAL, possibly a different species from *T. natans*) and in MNHN SAS 734 (an isolated mandible). These specimens come from the same locality as the holotype and are basal Pliocene in age. Scale bar, 5 cm.

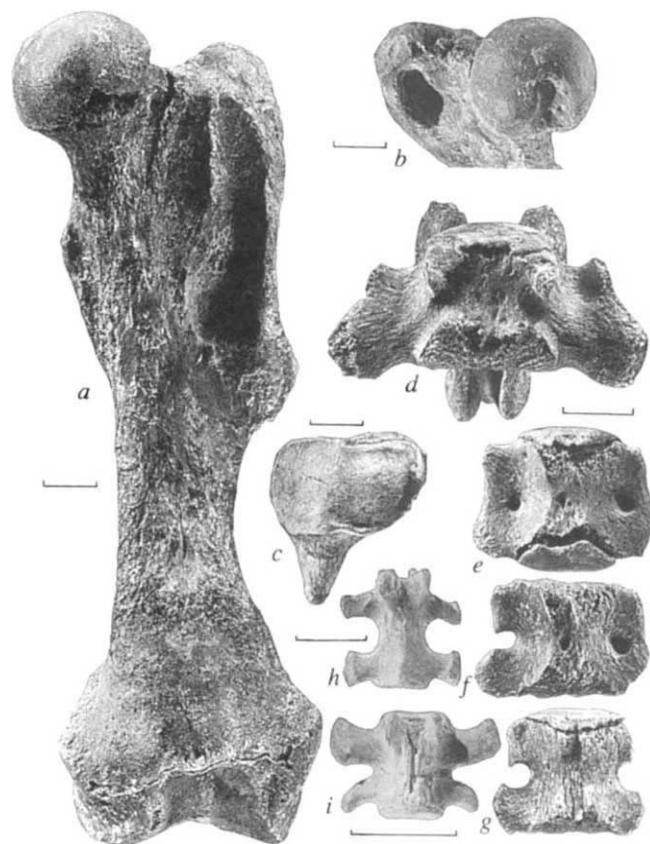


FIG. 2 *Thalassocnus natans* gen. et sp. nov. a, Left femur in anterior view (MNHN SAS 40); b, posterodistomedial view of the head of the same specimen showing the great width of the notch for the round ligament (ligamentum capitis femoris); c, anterior view of the patella (MNHN SAS 50); d, Anterior caudal vertebra (MNHN SAS 739) in ventral view; e, f, g, three caudal vertebrae from the middle third of the tail in ventral view (MNHN SAS 49). h, *Pteronura brasiliensis* twelfth caudal vertebra in dorsal view (MNHN AC 1870-99). i, *Castor fiber* eleventh caudal vertebra in dorsal view (MNHN AC 1965-22). Scale bars, 2 cm.

TABLE 1 Comparison of terrestrial/aquatic sloth limb bone proportions to those of terrestrial/aquatic rodents

	H	R	F	T	
<i>Thalassocnus</i>	243 (a.6)	249 (a.5)	289 (a.6)	246 (a.6)	
<i>Nothrotherium</i> UCMG 1020/04	256	239.3	231	178.5	
<i>Nothrotheriops</i> YPM 13198	387	340	341.5	253	
<i>Megalonyx</i> DMNH 25697, 25694, 25696, 25698	565	565	617	400	
<i>Megalocnus</i>	285 (a.5)	220 (a.2)	286 (a.4)	210 (a.2)	
<i>Eremotherium</i>	795	742 (a.2)	711 (a.4)	561 (a.3)	
<i>Castor</i>	83.6 (a.3)	83.2 (a.3)	105.6 (a.3)	126.4 (a.3)	
<i>Ondatra</i>	38.7 (a.2)	40.8 (a.2)	47.0 (a.2)	64.9 (a.2)	
<i>Neofiber</i>	26.8	28.1	33.5	40.5	
<i>Myocastor</i>	67.4 (a.2)	72.6 (a.2)	86.6 (a.2)	105.2 (a.2)	
<i>Geomys</i>	25.1	23.5	30.0	29.6	
<i>Marmota</i>	66.7	58.0	79.8	77.5	
<i>Spermophilus</i>	36.0	31.0	45.0	47.5	
<i>Cavia</i>	38.6	31.8	45.9	47.5	
<i>Erethizon</i>	79.1	69.6	93.9	81.7	
	R/H	H/F	T/F	R/T	H+R/F+T
<i>Thalassocnus</i>	0.98	0.86	0.85	0.99	0.92
<i>Nothrotherium</i>	0.93	1.11	0.77	1.34	1.21
<i>Nothrotheriops</i>	0.88	1.13	0.74	1.34	1.22
<i>Megalonyx</i>	1.00	0.92	0.65	1.41	1.11
<i>Megalocnus</i>	0.77	1.00	0.73	1.05	1.02
<i>Eremotherium</i>	0.93	1.12	0.79	1.32	1.21
Ave. TS	0.90	1.06	0.74	1.29	1.15
<i>Castor</i>	0.99	0.78	1.19	0.66	0.72
<i>Ondatra</i>	1.05	0.82	1.38	0.63	0.71
<i>Neofiber</i>	1.04	0.80	1.21	0.69	0.74
<i>Myocastor</i>	1.07	0.78	1.21	0.69	0.73
Ave. AR	1.04	0.79	1.25	0.67	0.72
<i>Geomys</i>	0.94	0.83	0.99	0.79	0.82
<i>Marmota</i>	0.87	0.84	0.97	0.74	0.79
<i>Spermophilus</i>	0.86	0.80	1.05	0.65	0.83
<i>Cavia</i>	0.82	0.84	1.03	0.67	0.75
<i>Erethizon</i>	0.88	0.84	0.87	0.85	0.85
Ave. TR	0.87	0.83	0.98	0.74	0.81

See legend to Fig. 3 for abbreviations.

distal process for the patellar ligament. This suggests a quadriceps femoris muscle smaller than in other gravigrades in which it was massive and functioned with the oblique position of the femur to permit an upright position for feeding²⁴. The buoyancy present during aquatic bottom feeding by *Thalassocnus* would have reduced the efforts of quadriceps femoris. The femoral neck is more distinct than in other nothrotheres. The fovea for the ligamentum capitis femoris (Fig. 2b) extends to the edge of the articular surface of the head of the femur where it forms a deep indentation. This condition is absent in other nothrotheres and indicates a very large ligament which suggests a potential for

abundant and strong abductions at the femoroacetabular joint. The third trochanter is shifted proximally and is continuous with the proximally expanded greater trochanter (both trochanters receive the insertions of the gluteal muscles). The morphology of the proximal extremity of the femur therefore indicates greater mobility and more powerful movements at the hip joint in *Thalassocnus* than in other nothrotheres²⁵. This feature, together with a tibia proportionally longer relative to the femur (a classical aquatic feature) than in other ground sloths (Fig. 3, Table 1), is consistent with a shift in the function of the hind limb towards aquatic or semi-aquatic life.

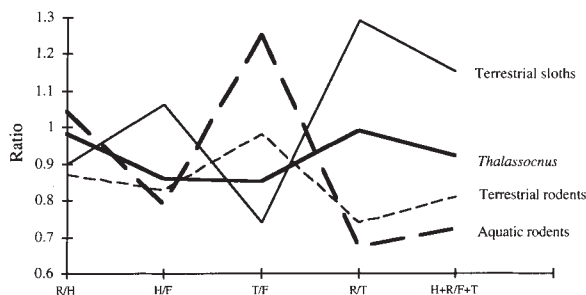


FIG. 3 Diagram comparing relationships of terrestrial sloth/aquatic sloth limb bone proportions to those of terrestrial rodents/aquatic rodents. Rodents have been chosen for comparisons because (1) they represent a group of mammals which includes both terrestrial and semi-aquatic forms, and (2) their limb proportions best compare with the sloths listed in Table 1. Each data point (except that of *Thalassocnus*) represents the average of the genera mentioned in Table 1. For each ratio considered, *Thalassocnus* has the same position relative to other

ground sloths as aquatic rodents to terrestrial rodents. It is noteworthy that a similar relationship of limb proportions exists between aquatic (otters and *Enaliarctos*) carnivores and *Thalassocnus* except for the R/H ratio (see below) which is between 0.7 and 0.8 and, therefore, closer to the terrestrial sloths, rodents, and some other carnivores (*Martes*, *Gulo*, *Felis*, *Ursus*). The comparison of the limb ratios of *Thalassocnus* to those of other ground sloths and to those of aquatic and terrestrial rodents therefore suggests a trend in the Peruvian sloth to modify the ancestral terrestrial nothrothere limb proportions towards aquatic limb proportions and is consistent with our interpretation of the biology of this animal as suggested by the taphonomy. The T/F ratio especially indicates that the tibia is proportionally longer in *Thalassocnus* than in the other sloths, although the tibia is not longer than the femur as in aquatic carnivores and rodents (where the tibia is always proportionally longer compared to the femur than in the terrestrial representatives). Comparisons with highly aquatic mammals (pinnipeds, sirenians), which have strongly modified limbs, is regarded here as irrelevant. H, Humerus; R, radius; F, femur; T, tibia, a(x), average of x number of specimens; Ave, average; TS, terrestrial sloth; TR terrestrial rodent; AR, aquatic rodent; DMNH, Dayton Museum of Natural History; UCMG, Universidade Catolica de Minas Gerais; YPM, Yale Peabody Museum.

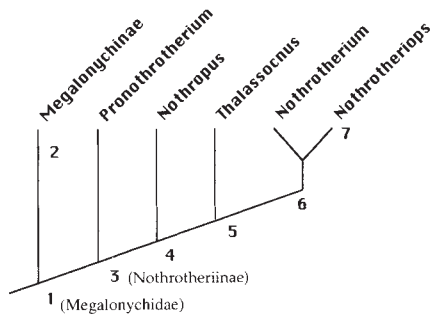


FIG. 4 Phylogenetic relationships of the Nothrotheriinae. The Nothrotheriinae are included within the Megalonychidae^{27,28} in contrast to other interpretations, which relate them to the Megatheriidae^{12,29}. Node 1, Megalonychidae: anterior tooth modified into caniniform and separated from cheek teeth by diastema, tuber calcis of calcaneum medio-laterally expanded, fifth metatarsal trapezoidal in shape with facets for cuboid and fourth metatarsal at oblique angle, third trochanter present on femur. Node 2, Megalonychinae: caniniform positioned at anterior margin of maxilla, cheek teeth triangular and bilophodont with transverse lophs convergent, medial trochlea of astragalus not modified into odontoid process, patellar surface of femur continuous with condylar surfaces. Node 3, Nothrotheriinae: caniniform positioned midway between anterior margin of maxilla and tooth row, cheek teeth quadrangular and bilophodont with transverse lophs parallel, longitudinal groove present on labial side of cheek teeth, post-palatine notch extends to last upper cheek tooth, supraorbital foramen present, predental spout of mandible elongated, ventral margin of predental spout straight, patellar groove of femur separated from condylar surfaces, astragalus with medial trochlea modified into odontoid process, proximal and second phalanges digit three pes co-ossified (unknown in *Nothropus*). Node 4, slight inflation of base of pterygoids. Node 5, loss of caniniform, post-palatine notch of palate positioned posterior to last upper cheek tooth, ventral margin of predental spout sigmoid relative to horizontal ramus, ungual phalanx of digit two of the manus dorsoventrally compressed. Node 6, pterygoids inflated with large pterygoid sinuses, ungual phalanx of digit one of the pes lost (unknown in *Thalassocnus*). Node 7, third trochanter forms continuous ridge with lateral epicondyle of femur.

The caudal vertebrae (Fig. 2d–g) of *Thalassocnus* have short but anteroposteriorly elongated transverse processes. In the first posterior caudal vertebrae, the transverse processes acquire a foramen for the passage of the segmental arteries²⁶. In the more posterior caudals, the foramen opens laterally and the transverse process becomes bifurcated. This condition is absent in other sloths but is observed in otters (Fig. 2h) and beavers (Fig. 2i), semi-aquatic mammals capable of powerful dorsoventral strokes of the tail.

Thalassocnus therefore indicates another pathway by which a terrestrial group has shifted to an aquatic habitat. The anatomical features are different from those in other semi-aquatic mammals (hippos, desmostylians, aquatic rodents, aquatic carnivores) because of the inherent restriction due to the anatomical constraints of being a sloth. In fact there seems to be many ways in which a mammal can become aquatic or semi-aquatic.

Thalassocnus is a nothrotheriine megalonychid (Fig. 4) which was living on the beaches of the Pliocene desert coast of the southern Peru. We suggest that it had semi-aquatic habits and was entering the near-shore waters (femur and patella morphology) where it grazed on sea-grasses or seaweeds (premaxilla morphology). Strong dorsoventral movements of the tail (morphology of the caudal vertebrae) would have helped to maintain a head-down feeding position. Swimming could have been by paddling of the hind limbs (morphology of femur head and length of tibia). *Thalassocnus* was certainly not a fast swimmer, but the taphonomy strongly supports semi-

aquatic habits and the anatomy is consistent with this interpretation. □

Received 1 February; accepted 22 March 1995.

- Marshall, L. G. & Cifelli, R. L. *Palaeovertebrata* **19**, 169–210 (1990).
- Webb, D. in *The Great American Biotic Interchange* (eds Stehli, G. & Webb, D.) 357–386 (Plenum, New York, 1985).
- White, J. J. *vert. Paleont.* **13**, 230–242 (1993).
- Hoffstetter, R. C. *r. hebd. Séanc. Acad. Sci., Paris* **D267**, 1273–1276 (1968).
- Muizon, C. de *Ed. Rech. gr. Civilisations, mém.* **8**, 1–150 (Trav. Inst. Fr. Et. Andines 22) (1981).
- Muizon, C. de & DeVries, T. J. *Geol. Rdsch.* **74**, 547–563 (1985).
- Alpers, C. N. & Brimhall, C. H. *Geol. Soc. Am. Bull.* **100**, 1640–1651, (1988).
- Sébrier, M., Marocco, R. & Macharé, J. in *Montagnes et Piedmonts* 49–69 (Revue Géographique des Pyrénées et du Sud-Ouest, Spec. issue, Toulouse, 1984).
- Muizon, C. de & Bellon, H. C. *r. hebd. Séanc. Acad. Sci., Paris* **D290**, 1063–1066 (1981).
- Muizon, C. de & Bellon, H. C. *r. hebd. Séanc. Acad. Sci., Paris* **303**, 1401–1404 (1986).
- Cartelle, C. & Fonseca, J. S. *Lundiana* **2**, 127–181 (1983).
- Paula Couto, C. *Tratado de Paleomastozoología* (Acad. Bras. Ci., Rio de Janeiro, 1979).
- Frailey, C. D. *Contr. Sci. Nat. Hist. Mus. L. A. Co.* **374**, 1–46 (1986).
- Stock, C. *Publs Carnegie Instn* **331**, 1–206 (1925).
- Muizon, C. de *Ed. Rech. Civilisations, mém.* **50**, 1–188 (Trav. Inst. Fr. Et. Andines 27) (1984).
- Muizon, C. de *Ed. Rech. Civilisations mém.* **78**, 1–244 (Trav. Inst. Fr. Et. Andines 42) (1988).
- Muizon, C. de *Nature* **365**, 745–748 (1993).
- Marocco, R. & Muizon, C. de *Bull. Inst. Fr. Et. Andines* **27**, 105–117 (1988).
- Cartelle, C. & Bohorquez, G. A. *Iheringia* **11**, 9–14 (1986).
- Solounias, N., Teaford, M. & Walker, A. *Paleobiology* **14**, 287–300 (1988).
- Hansen, R. M. *Paleobiology* **4**, 302–319 (1978).
- Naples, V. *Contr. Sci. Nat. Hist. Mus. L. A. Co.* **389**, 1–21 (1987).
- Paula Couto, C. *Annlis. Acad. brasil. Cienc.* **43**(suppl.), 499–513 (1971).
- Owen, R. *Description of the Skeleton of an Extinct Gigantic Sloth, Mylodon robustus, Owen, with Observations on the Osteology, Natural Affinities, and Probable Habits of the Megatheroid Quadrupeds in General* (R. & J. Taylor, London, 1842).
- White, J. L. J. *vert. Paleont.* **13**, 230–242 (1993).
- Miller, M. E., Christensen, G. C. & Evans, H. E. *Anatomy of the Dog* (Saunders, Philadelphia, 1964).
- Hoffstetter, R. *Traité de Paléont.* **VI**, 522–636 (Masson, Paris, 1958).
- Mones, A. *Cour. Forsch.-Inst. Senckenberg* **82**, 1–625 (1986).
- Patterson, B., Segall, W., Turnbull, W. D. & Gaudin, T. J. *Fieldiana Geol.* **24**, 1–79 (1992).

ACKNOWLEDGEMENTS. The research was funded by the Institut Français d'Études Andines (Lima, Peru) and the Muséum national d'Histoire naturelle (Paris, France). We thank D. Frey (Staatliches Museum für Naturkunde, Karlsruhe, Germany), K. Seymour (Royal Ontario Museum, Toronto) and B. Senut (Muséum national d'Histoire naturelle, Paris) who kindly gave access to the specimens under their care. Valuable insights were provided by G. F. Engelmann and C. E. Ray. Photographs are by D. Serrette (CNRS, Muséum national d'Histoire naturelle, Paris).

Experimentally induced transitions in the dynamic behaviour of insect populations

R. F. Costantino*, J. M. Cushing†, Brian Dennis‡ & Robert A. Desharnais§

* Department of Zoology, University of Rhode Island, Kingston, Rhode Island 02881, USA

† Department of Mathematics, University of Arizona, Tucson, Arizona 85721, USA

‡ Department of Fish and Wildlife Resources, University of Idaho, Moscow, Idaho 83844, USA

§ Department of Biology, California State University, Los Angeles, California 90032, USA

SIMPLE nonlinear models can generate fixed points, periodic cycles and aperiodic oscillations in population abundance without any external environmental variation. Another familiar theoretical result is that shifts in demographic parameters (such as survival or fecundity) can move a population from one of these behaviours to another^{1–4}. Unfortunately, empirical evidence to support these theoretical possibilities is scarce^{5–15}. We report here a joint theoretical and experimental study to test the hypothesis that changes in demographic parameters cause predictable changes in the nature of population fluctuations. Specifically, we developed a simple

model describing population growth in the flour beetle *Tribolium*¹⁶. We then predicted, using standard mathematical techniques to analyse the model, that changes in adult mortality would produce substantial shifts in population dynamic behaviour. Finally, by experimentally manipulating the adult mortality rate we observed changes in the dynamics from stable fixed points to periodic cycles to aperiodic oscillations that corresponded to the transitions forecast by the mathematical model.

We modelled the relationship linking larval, pupal and adult numbers at time $t+1$ to the number of animals at time t in the flour beetle (see ref. 9). The model is a system of three difference equations:

$$\begin{aligned} L_{t+1} &= bA_t \exp(-c_{ea}A_t - c_{el}L_t) \\ P_{t+1} &= L_t(1 - \mu_l) \\ A_{t+1} &= P_t \exp(-c_{pa}A_t) + A_t(1 - \mu_a) \end{aligned} \quad (1)$$

Here, L_t is the number of feeding larvae, P_t is the number of non-feeding larvae, and pupae and callow adults, and A_t is the

number of mature adults, at time t ; the unit of time (2 weeks) is taken to be the feeding larval maturation interval, so that after one unit of time a larva either dies or survives and pupates. This unit of time is also the cumulative time spent as a non-feeding larva, pupa and callow adult. The quantity $b > 0$ is the number of larval recruits per adult per unit of time in the absence of cannibalism. The fractions μ_l and μ_a are the larval and adult probabilities, respectively, of dying from causes other than cannibalism. The exponential nonlinearities account for the cannibalism of eggs by both larvae and adults and the cannibalism of pupae by adults. The fractions $\exp(-c_{ea}A_t)$ and $\exp(-c_{el}L_t)$ are the probabilities that an egg is not eaten in the presence of A_t adults and L_t larvae. The fraction $\exp(-c_{pa}A_t)$ is the survival probability of a pupa in the presence of A_t adults.

For fitting to time-series data, the model was converted to a stochastic model with noise added on a logarithmic scale:

$$\begin{aligned} L_{t+1} &= bA_t \exp(-c_{ea}A_t - c_{el}L_t + E_{1t}) \\ P_{t+1} &= L_t(1 - \mu_l) \exp(E_{2t}) \\ A_{t+1} &= [P_t \exp(-c_{pa}A_t) + A_t(1 - \mu_a)] \exp(E_{3t}) \end{aligned} \quad (2)$$

Here E_{1t} , E_{2t} and E_{3t} are random noise variables assumed to have a joint multivariate normal distribution with means of zero.

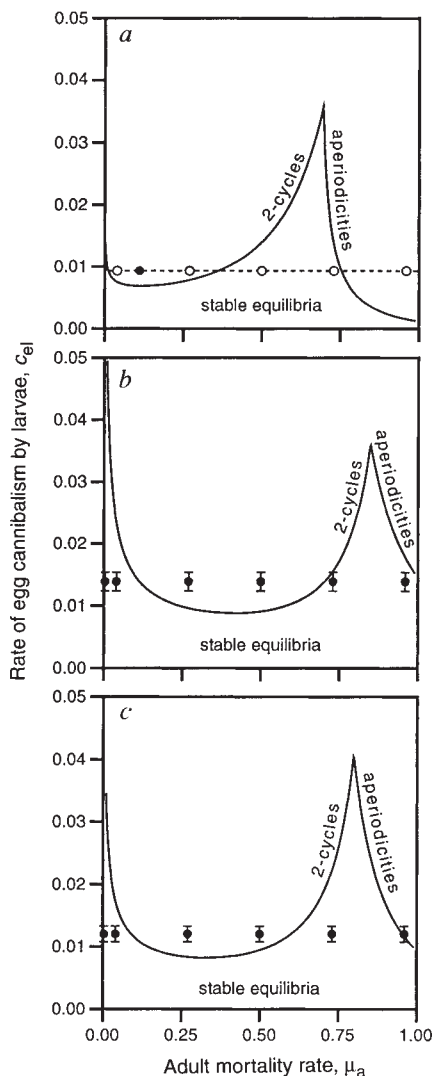


FIG. 1 Stability boundaries for the deterministic model (1) for three experiments. a, Sensitive strain¹⁶ with $b=11.68$, $c_{ea}=0.011$, $c_{pa}=0.017$, $\mu_l=0.513$. The filled circle locates the population in parameter space. The broken line and open circles indicate the extrapolated predicted dynamic behaviour as a function of adult mortality. b, RR strain with $b=7.88$, $c_{ea}=0.011$, $c_{pa}=0.004$, $\mu_l=0.161$. c, SS strain with $b=7.48$, $c_{ea}=0.009$, $c_{pa}=0.004$, $\mu_l=0.267$. The filled circles locate the experimental populations in parameter space. In b and c the bar represents a 95% confidence interval for c_{el} based on the profile likelihood.

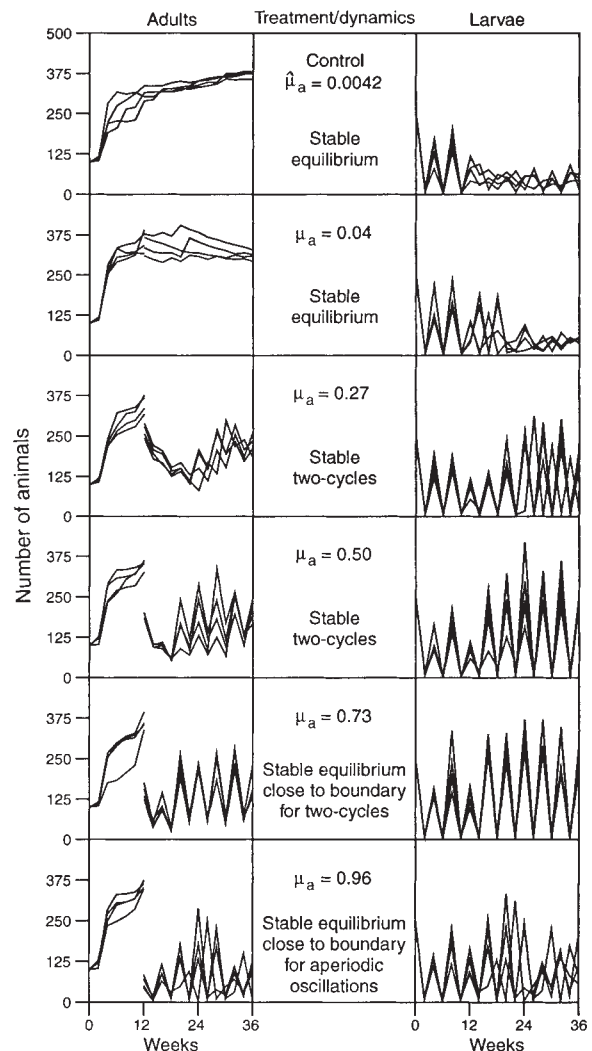


FIG. 2 Time-series data for individual replicates of the RR strain. At week 12 the experimental manipulation of adult mortality rate began. The centre column lists the imposed adult mortality rates and the location of the cultures in parameter space.

The noise variables represent the unpredictable departures of the observations from the deterministic model (1) due to environmental and other causes. Maximum-likelihood estimates of model parameters were calculated under the assumptions that the noise variables are correlated with each other but uncorrelated through time¹⁶. These assumptions were evaluated for all the data sets by standard diagnostic analyses of time-series residuals¹⁷.

In the study reported here we experimentally set adult mortality rates at values suggested by the analysis of historical time-series data¹⁶ (Fig. 1a) to place the experimental cultures in regions of different asymptotic dynamics: $\mu_a = 0.04, 0.27, 0.50, 0.73$ and 0.96 . There were also control cultures which were not manipulated. Cultures of *T. castaneum* (24 for each of two genetic strains, RR and SS) were initiated with 100 young adults, 5 pupae and 250 small larvae. Each population was contained in a half-pint (237 ml) milk bottle with 20 g of standard media and kept in a dark incubator at 31 °C. Every two weeks the L_t , P_t , and A_t stages were censused and returned to fresh media. This procedure was continued for 36 weeks. At week 12, four populations of each genetic strain were randomly assigned to each of the six treatments, and the imposition of adult mortalities began. Adult mortality was manipulated by removing or adding

adults at the time of a census to make the total number of adults that died during the interval consistent with the treatment value of μ_a . To counter the possibility of genetic changes in life-history characteristics, beginning at week 12 and continuing every month thereafter, the adults returned to the populations after the census were obtained from separate stock cultures maintained under standard laboratory conditions.

For each genetic strain the four replicates of each treatment were randomly assigned into two groups. With half of the data, we fitted model (2) to the time series of the RR (Fig. 2) and SS (Fig. 3) strains. The other half of the data were used for evaluating the model predictions. Analyses of time-series residuals¹⁷ indicated that the stochastic model described the data quite well. Using the model and maximum-likelihood parameter estimates, based on the time series for weeks 12 to 36, we calculated the stability boundaries (Fig. 1b, c) and bifurcation diagrams (Fig. 4) for each genetic strain.

The parameter estimates placed the control and $\mu_a = 0.04$ treatments in the region of stable equilibria. In the $\mu_a = 0.27$ treatment there was a transition in the dynamics: regular, albeit small, fluctuations were found in the adult data, whereas large and sustained oscillations were noted in larval numbers. These populations were located in the two-cycle zone. At $\mu_a = 0.50$ both strains were in the two-cycle region and displayed sustained oscillations, most noticeably in larval numbers. With $\mu_a = 0.73$ the RR strain was close to the two-cycle boundary but the SS strain underwent another transition and was clearly in the region of stable equilibria. Fluctuations in the RR strain were sustained (Fig. 2) whereas the fluctuations in the SS strain appeared to dampen (Fig. 3).

In the $\mu_a = 0.96$ treatment both strains underwent another transition in the dynamics. The RR strain was in the stable equilibria region and the SS strain was in the region of aperiodic oscillations; however, both were close to the boundary at which a bifurcation to aperiodicities occurs (Fig. 1b, c). Being close to the aperiodic bifurcation boundary means that the transients of

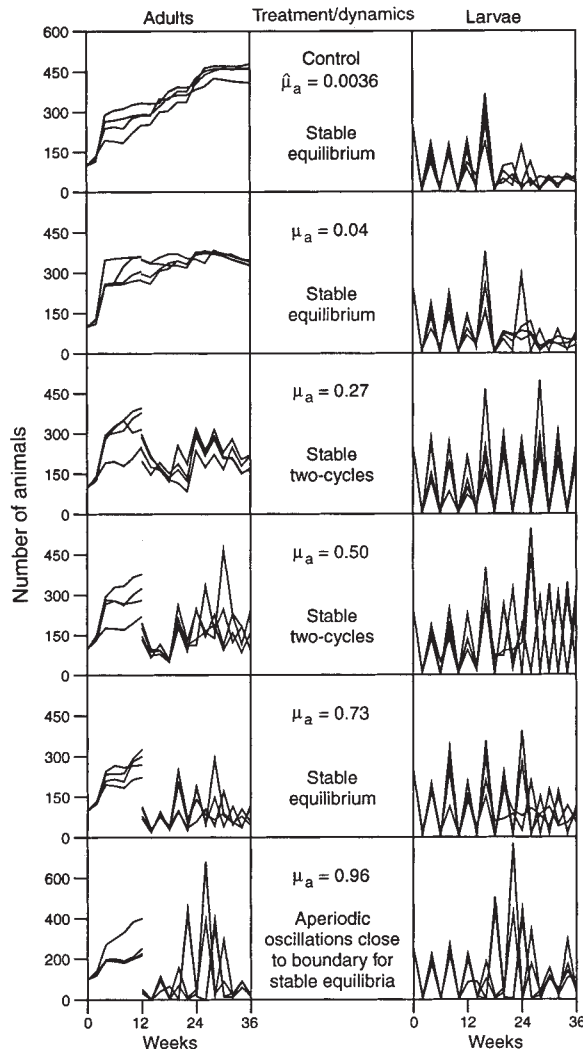


FIG. 3 Time-series data for individual replicates of the SS strain. At week 12 the experimental manipulation of adult mortality began. The centre column lists the imposed adult mortality rates and the location of the cultures in parameter space.

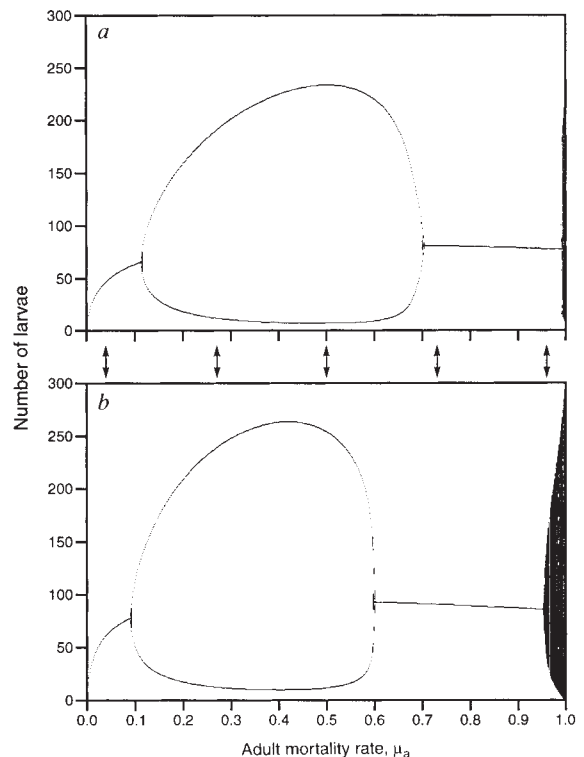


FIG. 4 Bifurcation diagrams for the model equations (1) for parameter values based on the experimental data. a, RR strain; b, SS strain. The double arrows indicate the adult mortality rate treatments.

the RR strain are expected to appear aperiodic and only slowly decay into the equilibrium. Both genetic strains displayed the expected aperiodic fluctuations (Figs 2 and 3).

Oscillatory approaches to point equilibria, stable periodic oscillations and aperiodic oscillations are difficult to distinguish in short time series with model-free methods. Our model-based parametric approach to classifying dynamic behaviour is statistically more powerful than non-parametric regression methods¹⁸ or parametric flexible-response-surface methods¹⁹. However, our classifications depend on the adequacy of the model (2). We minimized the risk of error by basing the model on detailed knowledge of a well-studied system⁹ and by thorough diagnostic analyses of the resulting time-series residuals¹⁶. The data (Figs 2 and 3) also give a visual impression of dynamic behaviours consistent with our classifications. Other, less-studied systems require statistical approaches that are more robust to variations in model form²⁰.

Our joint theoretical and experimental study has documented the transitions in the asymptotic dynamics of laboratory cultures of *Tribolium*. This rigorous verification of the predicted shifts in dynamical behaviour provides convincing evidence for the relevance of nonlinear mathematics in population biology. □

Received 21 December 1994; accepted 30 March 1995.

- May, R. M. *Science* **186**, 645–647 (1974).
- May, R. M. *J. theor. Biol.* **51**, 511–524 (1975).
- May, R. M. *Nature* **261**, 459–467 (1976).
- May, R. M. & Oster, G. F. *Am. Nat.* **110**, 573–599 (1976).
- Strong, D. R. in *Ecological Theory and Integrated Pest Management* (ed. Kogan, M.) 37–58 (Wiley, New York, 1986).
- Kareiva, P. in *Perspectives in Ecological Theory* (ed. Roughgarden, J., May, R. M. & Levin, S. A.) 68–88 (Princeton University Press, Princeton, NJ, 1989).
- Bartlett, M. S. *J. R. statist. Soc. A52*, 321–347 (1990).
- Logan, J. A. & Hain, F. (eds) *Chaos and Insect Ecology* (Virginia Polytechnic Institute and State University, Blacksburg, VA, 1991).
- Costantino, R. F. & Desharnais, R. A. *Population Dynamics and the Tribolium Model: Genetics and Demography* (Springer, New York, 1991).
- Logan, J. A. & Allen, J. C. *A. Rev. Ent.* **37**, 455–477 (1992).
- Hastings, A., Hom, C. L., Ellner, S., Turchin, P. & Godfray, H. C. *J. A. Rev. Ecol. Syst.* **24**, 1–33 (1993).
- Murdock, W. W. & McCauley, E. *Nature* **316**, 628–630 (1985).
- Grenfell, B. J., Price, D. J., Albon, S. D. & Clutton-Brock, T. H. *Nature* **355**, 823–826 (1992).
- Hanski, I., Turchin, P., Korpiimäki, E. & Henttonen, H. *Nature* **364**, 232–235 (1993).
- Tilman, D. & Wedin, D. *Nature* **353**, 653–655 (1991).
- Dennis, B., Desharnais, R. A., Cushing, J. M. & Costantino, R. F. *Ecol. Monogr.* (in the press).
- Tong, H. *Non-linear Time Series: a Dynamical System Approach* (Oxford Univ. Press, Oxford, 1990).
- McCaffrey, D. J., Ellner, S., Gallant, A. R. & Nychka, D. W. *J. Am. statist. Ass.* **87**, 682–695 (1992).
- Turchin, P. & Taylor, A. *Ecology* **73**, 289–305 (1992).
- Turchin, P. *Oikos* **68**, 167–172 (1993).

ACKNOWLEDGEMENTS. This work was supported by the NSF.

Interactions in the flexible orientation system of a migratory bird

Kenneth P. Able & Mary A. Able

Department of Biological Sciences, State University of New York, Albany, New York 12222, USA

MIGRATING birds rely on interacting compass senses: magnetic, star, polarized light and perhaps Sun compasses^{1,2}. During the development of orientation mechanisms, celestial rotation of stars at night³ and of polarized skylight patterns during the day time⁴ provide information about true compass directions that calibrates the direction of migration selected using the magnetic compass^{3,11}. It might often be advantageous to adjust the magnetic preference by a geographic reference, especially at high northern latitudes where magnetic declination is large. Paradoxically, a magnetic preference so calibrated will be reliable only within a region of similar declination unless magnetic orientation remains open to calibration in older birds, something that earlier studies suggested was not the case^{1,2}. We report here that in the Savannah sparrow (*Passerculus sandwichensis*) the same sort of calibration of magnetic orientation found in very young birds also occurs in older individuals exposed during the migration period to clear day and night skies within a shifted magnetic field.

The Savannah sparrow is a medium-distance nocturnal migrant that nests in grassland, meadow and tundra across North America from the northern United States northwards to 60–70° N. It migrates to the southern United States southward to northern Central America in winter. Adult Savannah sparrows possess magnetic^{12,13} and star compasses⁸, and use visual cues associated with clear sunset skies (including skylight polarization patterns)^{2,14,15} for orientation. We captured Savannah sparrows in local breeding fields in early September, 1994, before autumn migration. Of the 39 birds used in the experiment, 11 were adults (older than 1 year) and 28 were born during the 1994 nesting season (immature) (age determined by degree of skull ossification). Except during experimental and control treatments, the birds were housed in individual cages in a normal

TABLE 1 Experimental treatments

Group	Order of tests	Field shift	Dates of exposure (1994)	
			C	S
(1) 4 adults, 6 immatures	C, S	90° CCW	9 Oct.–13 Oct.	18 Oct.–27 Oct.
(2) 3 adults, 6 immatures	S, C	90° CCW	27 Oct.–31 Oct.	13 Oct.–18 Oct.
(3) 4 adults, 6 immatures	C, S	90° CW	13 Oct.–18 Oct.	3 Nov.–15 Nov.
(4) 10 immatures	S, C	90° CW	15 Nov.–24 Nov.	27 Oct.–31 Oct.

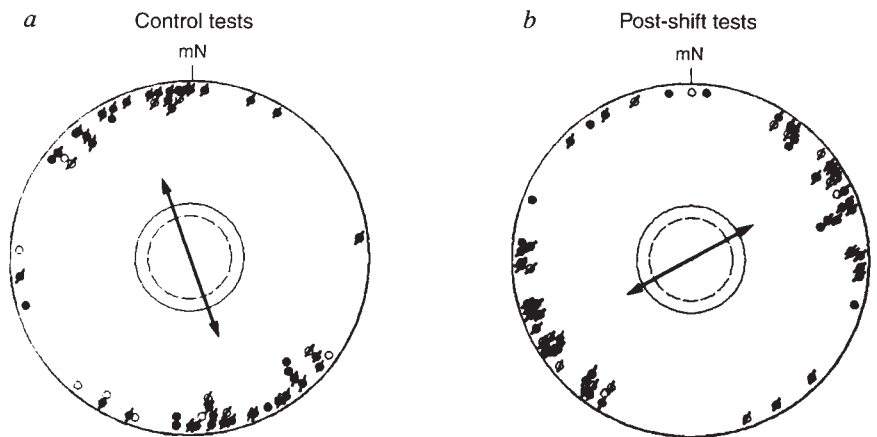
C, Outdoor exposure in unshifted magnetic field (control condition); S, outdoor exposure in shifted magnetic field (shifted field condition); CCW, counterclockwise; CW, clockwise.

magnetic field and under ambient photoperiod, but had no exposure to the outdoors or sky. Groups of 10 birds were placed in a single cage centred within a large (2.7 m) Rubens coil¹⁶ that was used to shift the magnetic field direction. The birds remained in the cage day and night until the group had experienced four clear days and four clear nights. They were then returned to individual cages and their magnetic orientation recorded in covered Emlen funnel orientation cages¹⁷ during tests conducted indoors. Each group of birds was exposed to the day and night sky within coils under two conditions: (1) in an unshifted magnetic field (5.5×10^4 nT; inclination, 70°; declination, 13° W) (control condition); and (2) in a magnetic field (5.6×10^4 nT; inclination, 64°) shifted either 90° clockwise or 90° anticlockwise (shift condition). Orientation was recorded from each individual following each type of exposure (see Table 1 for details).

Savannah sparrows typically show axially bimodal magnetic migration orientation when tested over several nights^{8,10,13}. Individual birds usually orient in a unimodal direction in each test, but often switch to more or less opposite directions on subsequent nights. This is true of hand-raised birds as well as wild-caught immatures and adults. Tested immediately after outdoor exposure in the control condition, both adults and immatures exhibited NNW–SSE magnetic orientation (Fig. 1a), typical of Savannah sparrows that have been raised outdoors in situations where magnetic and geographic directions were roughly coincident^{4,8,10,13}. After exposure under the shifted field condi-

FIG. 1 Magnetic orientation of the birds tested indoors in orientation cages covered with white translucent plastic sheets. *a*, Orientation of the birds after exposure to the clear daytime and night sky for four days in an unshifted magnetic field. These birds showed NNW–SSE magnetic orientation ($r=0.519$, modes= 160° and 340° , $P<0.001$, Rayleigh test, $n=37$ birds). *b*, Orientation of birds after exposure to the clear daytime and night sky for four days in a magnetic field shifted 90° clockwise (magnetic N equals geographic E) or 90° anticlockwise (magnetic N equals geographic W). These birds showed NE–SW magnetic orientation ($r=0.416$, modes= 61° and 241° , $P<0.001$, Rayleigh test, $n=38$ birds). Unfilled points represent the orientation of adult birds, solid points represent immatures. Points with a slash mark indicate hopping activity of birds that showed statistically bimodal orientation¹⁸. The broken and solid inner circles represent, respectively, the 5% and 1% significance levels for the Rayleigh test. mN, Magnetic north.

METHODS. Orientation tests were done in a wooden frame garage with an ambient magnetic field of total intensity of 5.4×10^4 nT, inclination 70° . Birds were prevented from seeing outside the test cages by the translucent covers. A dim incandescent light, passing through two diffusers and the translucent cage cover, provided a light of 0.2–0.4 lux within the cages. This variation in light intensity occurred across the



array of test cages. Light intensity within each cage was uniform. Tests began within 1 h of lights-off and lasted for 3 h. Each point represents the mean of the modal orientation directions of each bird (each bird tested 3–5 times). The orientation records were analysed blind with respect to treatment²⁸.

tion, the same birds oriented magnetic NE–SW (Fig. 1*b*). The distributions of orientation directions under the two test conditions were significantly different ($P<0.001$, Watson's U^2 test; no difference in dispersion, non-parametric test for dispersion¹⁸), and the difference in direction under the two test conditions (81° or 99° , depending upon sense of the shift) is very close to the 90° predicted by the hypothesis that magnetic orientation remains open to calibration by visual cues throughout life (predicted direction fell within 95% confidence intervals around the mean orientation direction).

Both adult (70%) and immature (54%) birds responded to the experimental treatment by shifting orientation direction and the proportion showing the predicted shifts in orientation did not differ between the two age classes. Tested after exposure in the shifted magnetic field, adult birds ($n=10$ individuals with orientation data under both conditions) exhibited a mean shift of 91° ($r=0.807$, $P<0.001$, Rayleigh test) relative to their individual means when tested after exposure under the control condition; immatures ($n=26$ individuals with orientation data under both conditions) exhibited a mean shift of 101° ($r=0.710$, $P<0.001$, Rayleigh test). Clockwise and anticlockwise field shifts produced similar changes in orientation (mean shift for clockwise groups was 89° , $r=0.798$, $P<0.001$; mean shift for anticlockwise groups was 110° , $r=0.687$, $P<0.001$, Rayleigh tests).

These results suggest that the interactions and relationships among orientation mechanisms are similar in young birds and older individuals. Exposure to large declination for several days can produce calibration of magnetic orientation in young birds before the first migration, in immatures during the first migration season, and in adults that have completed at least one round-trip migration. The importance of geographic compass directions in calibrating other orientation systems does not end before the first migration. The conventional view has developed that the hierarchical relationship among orientation cues changes as young birds reach the age of first migration^{1,2,5}. Whereas celestial rotation assessed by visual cues seemed to be of overriding importance during early development, cue-conflict experiments during migration suggested precedence of other cues. Our results reported here indicate that no real difference exists in the relationships of orientation mechanisms in younger versus older birds. The fact that in cue-conflict experiments of short duration

or involving stationary star patterns, the magnetic field took precedence over and could even assign directionality to stars^{19–23} is not necessarily inconsistent with calibration of magnetic orientation by several days of exposure to visual cues. The apparent differences may have resulted from the ways in which the experiments were done. Most cue-conflict experiments with birds of migratory age have lasted for only a few hours or at most a whole night, whereas in the ontogeny experiments birds were exposed to the cue conflict for many days. The design of the experiments may have precluded access by the birds to any information that could have revealed the axis of celestial rotation. It is therefore not surprising that magnetic orientation seemed stable and closed to modification in older birds.

The potential for repeated recalibration that we have described enables migratory birds to respond to spatial and temporal variability in the quality or availability of orientation information. Birds may encounter considerable variation in magnetic declination as they migrate and unfamiliar star patterns will be met as a first-time migrant moves southward. In the vicinity of the magnetic equator the inclination magnetic compass will face difficulties because field lines are horizontal, and birds that cross the magnetic equator must reverse their directional response to the field in order to continue southward^{24,25}. Visual orientation cues may be obscured by cloud cover and some, such as sunset position, will change direction over the migration season. This latter cue has been shown to be calibrated by polarized skylight patterns in adult birds²⁶. Contrary to the general consensus in the field³, plasticity in the migratory orientation system is not confined to a sensitive period in the early life of the animal. Only the initial calibration of star patterns by stellar rotation seems to be irreversible once the bird reaches migratory age²⁷. This is not surprising because under natural conditions the star compass should not require recalibration. □

Received 9 January; accepted 22 March 1995.

1. Wiltschko, W. & Wiltschko, R. in *Orientation in Birds* (ed. Berthold, P.) 16–37 (Birkhäuser, Basel, 1991).
2. Able, K. P. *Trends Ecol. Evol.* **8**, 367–371 (1993).
3. Able, K. P. & Able, M. A. *Nature* **347**, 378–380 (1990).
4. Able, K. P. & Able, M. A. *Nature* **364**, 523–525 (1993).
5. Able, K. P. in *Orientation in Birds* (ed. Berthold, P.) 166–179 (Birkhäuser, Basel, 1991).

6. Able, K. P. & Able, M. A. *J. comp. Physiol. A* (in the press).
7. Witschko, W., Däum, P., Fergenbauer-Kimmel, A. & Witschko, R. *Ethology* **74**, 285–292 (1987).
8. Bingman, V. P. *Behaviour* **87**, 43–53 (1983).
9. Bingman, V. P., Beck, W. & Witschko, W. in *Migration: Mechanisms and Adaptive Significance* (ed. Rankin, M. A.) 544–552 (Marine Sci. inst., Univ. Texas, Port Aransas, 1985).
10. Able, K. P. & Able, M. A. *Anim. Behav.* **39**, 905–913 (1990).
11. Prinz, K. & Witschko, W. *Anim. Behav.* **44**, 539–545 (1992).
12. Bingman, V. P. *Anim. Behav.* **29**, 962–963 (1981).
13. Able, K. P. & Able, M. A. *Ethology* **93**, 337–343 (1993).
14. Moore, F. R. *Anim. Behav.* **28**, 684–704 (1980).
15. Moore, F. R. *Anim. Behav.* **33**, 657–663 (1985).
16. Kirschvink, J. L. *Bioelectromagnetics* **13**, 401–411 (1992).
17. Emlen, S. T. & Emlen, J. T. *Auk* **83**, 361–367 (1966).
18. Batschelet, E. *Circular Statistics in Biology* (Academic, New York, 1981).
19. Witschko, W. & Witschko, R. *J. comp. Physiol. A* **109**, 91–99 (1976).
20. Witschko, W. & Witschko, R. *Zeits. f. Tierpsychol.* **37**, 337–355 (1975).
21. Witschko, W. & Witschko, R. *Zeits. f. Tierpsychol.* **39**, 265–282 (1975).
22. Bingman, V. P. *Auk* **104**, 523–525 (1987).
23. Beason, R. C. *J. Ornithol.* **128**, 317–324 (1987).
24. Witschko, W. & Witschko, R. *Ethology* **91**, 70–74 (1992).
25. Beason, R. C. *Ethology* **91**, 75–80 (1992).
26. Phillips, J. B. & Moore, F. R. *Behav. Ecol. Sociobiol.* **31**, 189–193 (1992).
27. Emlen, S. T. in *Animal Orientation and Navigation* (eds Galler, S. R., Schmidt-Koenig, K., Jacobs, G. J. & Belleville, R. E.) 191–210 (NASA, Washington DC, 1972).
28. Cherry, J. D. & Able, K. P. *Auk* **103**, 225–227 (1986).

ACKNOWLEDGEMENTS. We thank C. Walcott for use of the magnetometer, V. P. Bingman for comments, and the NSF for support of this work.

Head position signals used by parietal neurons to encode locations of visual stimuli

Peter R. Brothie*, Richard A. Andersen*†, Lawrence H. Snyder* & Sabrina J. Goodman

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

THE mechanism for object location in the environment, and the perception of the external world as stable when eyes, head and body are moved, have long been thought to be centred on the posterior parietal cortex^{1–8}. However, head position signals, and their integration with visual and eye position signals to form a representation of space referenced to the body, have never been examined in any area of the cortex. Here we show that the visual and saccadic activities of parietal neurons are strongly affected by head position. The eye and head position effects are equivalent for individual neurons, indicating that the modulation is a function of gaze direction, regardless of whether the eyes or head are used to direct gaze. These data are consistent with the idea that the posterior parietal cortex contains a distributed representation of space in body-centred coordinates.

We recorded from 128 cells in 2 monkeys trained in a task requiring saccades to visual targets from independent head and eye positions. Each monkey was trained to orient its head towards a large fixation spot (Fig. 1a and d) and its eyes to a small fixation spot (Fig. 1b and e). In each trial the monkey maintained fixation for a predefined period; a peripheral visual stimulus then appeared and the monkey made a saccadic eye movement to it (Fig. 1c and f). We analysed the neuron's activities during two time segments aligned to the beginning of the saccade. The first time segment was from 200 ms to 25 ms before the beginning of the saccade, and largely reflects activity related to the visual target, whereas the second time segment was from 25 ms before to 75 ms after saccade initiation, and largely reflects activity related to the eye movement⁹.

In the first experiment we examined the effect of head position on the receptive fields of 51 of the 128 neurons by mapping them at two different head positions, 16° to the left and 16° to the right of straight ahead. The monkey held its head in each of the two positions with an initial eye position straight ahead in the orbits. Recording the activity of the cell during the presentation of the stimulus provided a tuning curve of the cell's activities to eight different radial directions in each of the two head positions.

Figure 2a shows the tuning curves at the two head positions for one of the cells. This cell had a receptive field centred down and to the left (at an eccentricity of 16°) when the animal was fixating to the right. Because this particular recording was made from the right hemisphere, the left visual field is contralateral with respect to the recorded hemisphere, and the right visual field is ipsilateral. If the receptive field was referenced with respect to the body, rather than with respect to the head or eyes, fixating to the left should rotate the centre of this receptive field anti-clockwise (relative to the fixation point) by approximately 90° in the frontal plane to maintain the receptive field at the same location in space. Instead, the best direction was the same for the two head positions, but the magnitude of the activity at the peak was increased by 50% for contralateral (left) fixation.

We constructed a population tuning curve using all 51 cells, and investigated whether the curve shifted with changes in head position (Fig. 2b). The receptive field centres for the contralateral head position of all the cells were aligned at 0° on the abscissa, and the activity was averaged across the population of cells for both head positions. If the cells have receptive fields in body-centred coordinates, then the tuning curve for the ipsilateral head position should be shifted towards the contralateral side of the graph (to the right). However, the average tuning curve for the population peaks at 0° for both head positions. Instead, the magnitude of the population activity decreases for ipsilateral gaze. This decrease is due to the fact that the modulation by head position usually increases for the contralateral direction for area LIP neurons, which make up approximately two thirds of the sample. Therefore, a change in head position does not systematically change the peaks of the tuning curves of posterior parietal neurons. Similar effects were found for eye position in previous studies^{10–12}.

To test the effect of head position and eye position on the magnitude of the visual and saccade activities of cells at several gaze directions, we performed a second experiment in which we recorded the magnitude of the activities from each of five different head positions and five corresponding eye positions. After the receptive field of a cell was mapped and the preferred direction of the cell determined, the monkey was required to make saccades in this preferred direction from each of five different fixation points on the horizontal axis. These saccades were made under two conditions. First, the effect of head position was examined by having the monkey orient its head towards each fixation point with its eyes centred in their orbits (Fig. 1d–f). Second, the animal positioned its head straight forward with its eyes deviated to each of the fixation points (Fig. 1a–c). This allowed us to determine separately the modulation of the cell's activities due to eye and head position. Figure 3a and b shows the effects of head position and eye position, respectively, on the activity of a cell which preferred saccades directed contralaterally with 16° amplitude. In Fig. 3c the magnitude of the activity is plotted as a function of initial gaze angle. Two plots are shown, one for different head positions and one for different eye positions. Linear relationships between the activity of the cell and the initial gaze angle were demonstrated in both situations (as determined by linear regression and the analysis of variance of the residuals), with their slopes representing the 'gain fields' for eye and head position in this cell. Because the gain fields were very similar, it can be concluded that the activity of this cell was modulated by gaze position (the sum of eye and head position).

* Present addresses: Medical Imaging Department, St Vincent's Hospital, Fitzroy, Victoria 3065, Australia (P.R.B.); Division of Biology, 216-76, California Institute of Technology, Pasadena, California 91125, USA (R.A.A. and L.H.S.).

† To whom correspondence should be addressed.

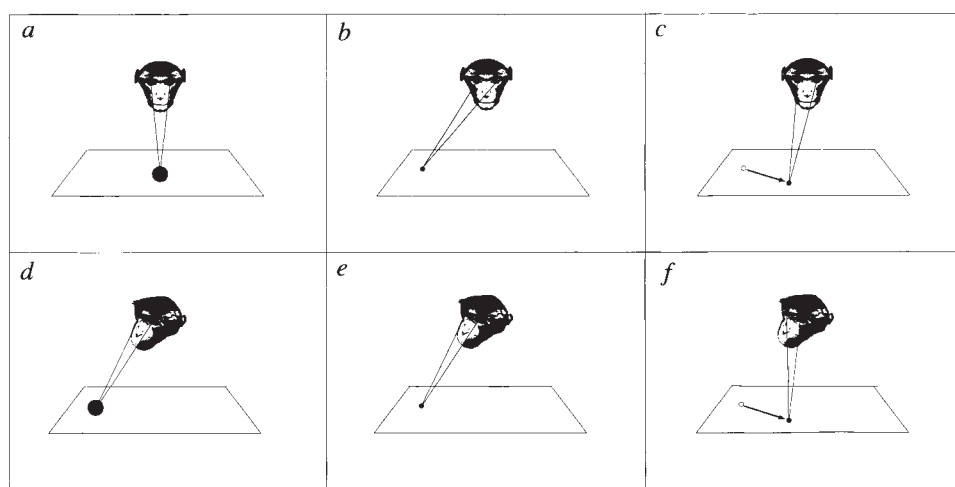
A significant modulation of activity by head position was found for 32 cells in the first time segment and 31 in the second, with 52 of 128 cells showing significant modulation in at least one of these two time segments ($P < 0.05$). The relationship between the gain fields for head and eye position for the two time segments is shown in Fig. 4. The slope of the head gain field (vertical axis) is plotted against the slope of the eye gain field (horizontal axis) for each neuron. If a cell displays identical gain fields for head and eye position its point will lie on a 45° positive diagonal in the graph. Most of the points for both time segments were clustered about this diagonal. Therefore, for most cells that demonstrated statistically significant gain fields for head position, the gain fields were best described as a function of gaze position (eye and head position).

These results show that the visual receptive fields of posterior parietal neurons are in retinal (eye-centred) coordinates, but are

strongly modulated by gaze direction. Thus individual cells carry information about retinal position, eye position and head position. However, each individual cell is ambiguous about the location of a target in body-centred coordinates because its activity may vary with any one of these three parameters. Thus a particular rate of firing of a single cell can correspond to many locations in space. Interestingly, when we trained neural networks to convert retinal receptive fields to body-centred locations, the units that perform this coordinate transformation develop gain fields for eye and head position which are also a function of gaze direction, similar to those found in the experimental data¹³. Thus the current results are consistent with a model of the posterior parietal cortex in which space is represented in body-centred coordinates, not at the level of single cells, but at the level of populations of cells. This distributed representation could be the final stage for coding locations in space, or it could be used as

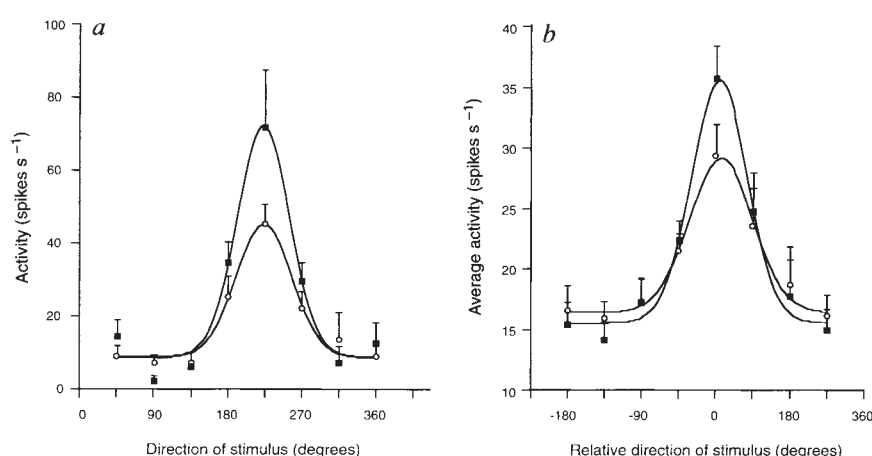
FIG. 1 The monkey uses its eyes (a–c) or head (d–f) to change the initial gaze direction before a saccade. a, d, Each trial begins with a 0.4° diameter fixation point appearing on the tangent screen at eye level. The animal orients its eyes and head toward the stimulus. b, e, The 0.4° stimulus disappears and a smaller, 0.2° diameter stimulus appears at either a different or the same location. The monkey maintains its previous head position and, if necessary, deviates its eyes to fixate the smaller stimulus. c, f, After 1,500 ms the fixation light is turned off and another 0.2° stimulus appears at a peripheral location. The monkey saccades to the new location while keeping its head still, and the sensory and saccade activities are recorded. For cells with both visual and saccade activity, the visual receptive fields and saccadic motor fields overlay one another in oculocentric coordinates²⁸ and eye position has the same modulation effects on both¹⁰. Control experiments, using a delayed-saccade task which separates temporally the visual and saccade activities, determined that head position has a similar modulation effect on both visual and saccade activities for individual neurons.

METHODS. Experiments were performed in the dark. Gaze direction was



recorded using the scleral search coil technique, and head position was recorded using a high-precision potentiometer connected to the head post which allowed the animal to move its head horizontally. Receptive fields were mapped systematically in 8 directions at eccentricities of 8° and/or 16°. If the cell's receptive field eccentricity was reasonably close to one of these two distances, further tests were made to assess the effects of head and eye position.

FIG. 2 a, Tuning curves of one cell at two head positions, calculated from the first time segment (200 ms to 25 ms before saccade initiation). The horizontal axis indicates the direction of the stimulus, where 0° was directly to the right (ipsilateral), and 180° was to the left (contralateral), and 90° was directed up. Filled squares indicate contralateral head position, and open circles ipsilateral head position in both panels. Bars indicate one standard error. b, Plot showing the average tuning curve of the population of 51 cells, calculated from the first time segment. The maximal activity of each cell at the contralateral head position was aligned to 0° on the horizontal axis. The region to the left of 0° is for ipsilateral direction with respect to the receptive field centre of each cell, and to the right of 0° indicates contralateral directions. The vertical axis is mean activity of the population, with bars indicating one standard error.



an intermediate step in the construction of body-centred receptive fields elsewhere in the brain^{14,15}.

The saccade signals, like the visual signals, are modulated by eye and head position, suggesting that they are both coded in the same distributed, body-centred coordinate frame. That some cells show only sensory or only movement activity, whereas others show both, is consistent with a distributed coding strategy in which each cell is not required to carry all of the signals involved in a neural computation. The role of the saccade signal in the posterior parietal cortex is at present not clear. It could represent an efference copy signal or proprioceptive signal which updates the spatial representation with each eye movement¹⁶. The saccade activity of nearly all area 7a neurons and nearly

one-third of the LIP cells begins after the beginning of the eye movement, strongly suggesting a feedback role⁹. However, the many LIP cells that are activated well in advance of a saccade could code the decision or intention to make an eye movement. Activities related to intentions to make eye movements have been shown in area LIP by using delayed saccade tasks which separate in time the planning and triggering of saccades¹⁷.

Possible sources for the head position signal are proprioceptive inputs from the neck muscles^{19–22}, an integration of the vestibular signal^{23–26} or an integration of corollary discharge associated with head movement. The last of these seems unlikely to be the sole source because the signal has been observed when the head is rotated by the experimenter rather than by the animal

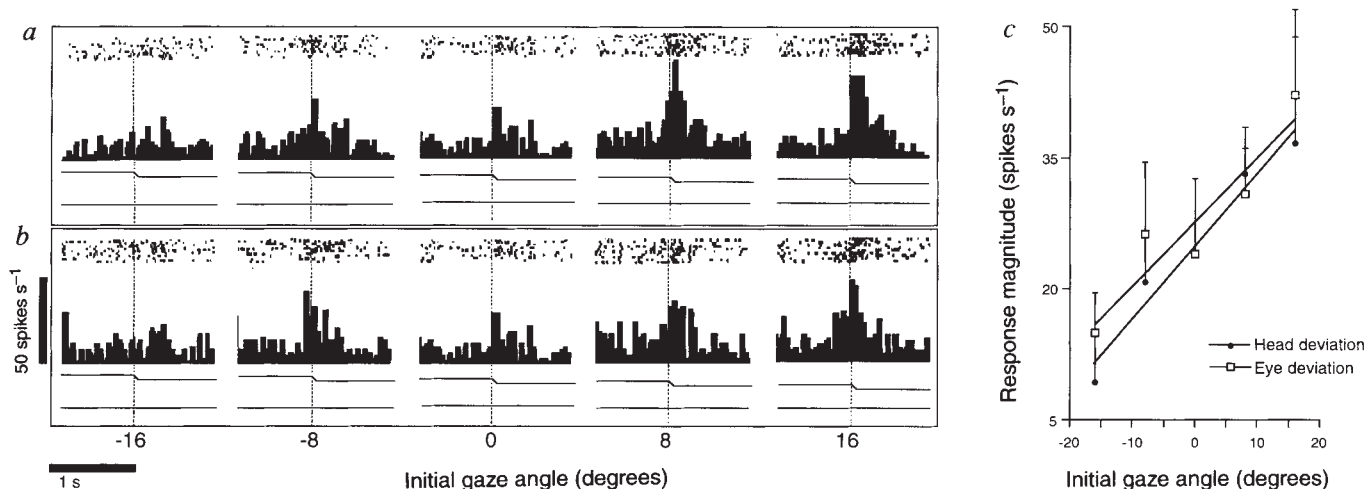
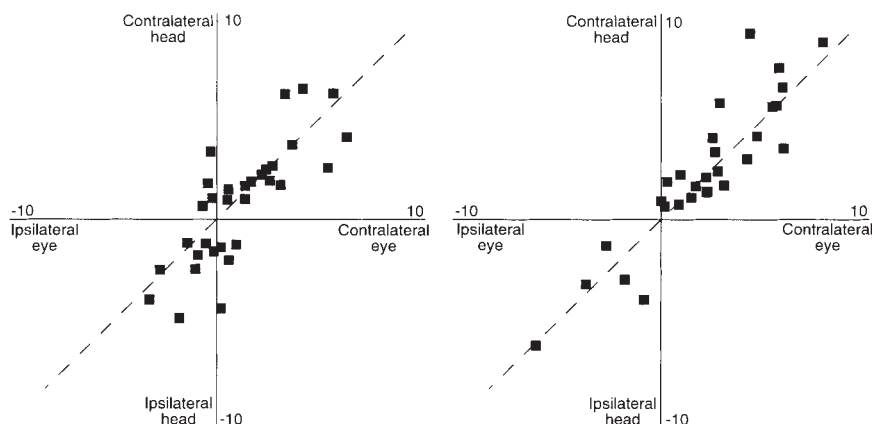


FIG. 3 Activity of a cell while the monkey is making identical saccades to the left from a fixation point placed at five different gaze positions on the horizontal plane. Gaze directions of $\pm 16^\circ$, $\pm 8^\circ$ and 0° directions were used in approximately two-thirds of the cells showing significant head effects, and $\pm 24^\circ$, $\pm 12^\circ$ and 0° were used for the remainder. All trials are aligned with the onset of the saccade, indicated by the vertical broken lines. *a*, The animal has its head oriented towards each of the fixation points, with the eyes centred in their orbits before each saccade. *b*, The head of the animal is directed towards the centre of the screen

(0°) with the eyes deviated towards each of the fixation points before each saccade. *c*, Magnitude of the cell's activity around the time of the saccade (25 ms before to 75 ms after onset of the saccade) as it varies with initial gaze position. Bars indicate one standard error. A linear relationship of activity with gaze direction was confirmed by a significant linear regression ($P=0.009$ for head, $P=0.023$ for eye) and by a non-significant analysis of variance (ANOVA) of the regression residuals ($P=0.992$ for head, $P=0.944$ for eye).

FIG. 4 Plot of the slopes of the regression lines for head and eye position for each cell with a significant modulation by head position ($P<0.05$). The vertical axis indicates the slope of the regression line for eye position. The horizontal axis shows the slope of the regression line for head position. Left, Plot for time segment 1 ($n=32$). A linear regression for these data points resulted in a line of slope 1.0 ± 0.1 s.e.m. and intercept of -0.3 ± 0.4 s.e.m. Right, Plot for time segment 2 ($n=31$) resulted in slope 1.0 ± 0.1 s.e.m. and intercept 0.1 ± 0.3 s.e.m. When computing the slopes of the gain fields, only 13% (4 of 32) of the time segment 1 cells and only 23% (7 of 31) of the time segment 2 cells had a significant departure from linearity ($P<0.05$, one-way ANOVA on residuals). Of the cells with no significant linear slope for head position ($P>0.05$), only 4% (3 of 76) had a significant nonlinearity in time segment 1, and 7% (5 of 76) in time segment 2 ($P<0.05$). A total of 32 cells were modulated by eye position in at least one of the time segments, but not head position in either time segment. Thus a large percentage (66%) of the



cells were modulated by eye or eye-and-head position. This percentage is substantial, especially considering the fact that we did not test for vertical eye or head position effects.

in a willed movement. Experiments currently in progress suggest that both neck proprioceptive signals and integrated vestibular signals contribute to the head position signal²⁷. □

Received 5 December 1994; accepted 24 February 1995.

1. Critchley, M. *The Parietal Lobes* (Hafner, New York, 1953).
2. Andersen, R. A. in *Handbook of Physiology: The Nervous System. Higher Functions of the Brain* Sect. 1, Vol. V, Ch. 12 (eds Plum, F., Mountcastle, V. B. & Geiger, S. R.) 483–518 (American Physiology Society, Bethesda, Maryland, 1987).
3. Mountcastle, V. B. *Neurosci. Res. Prog. Bull.* **14** (Suppl.), 1–47 (1976).
4. Lynch, J. C. *Behav. Brain Sci.* **3**, 485–534 (1980).
5. Sakata, H., Shibutani, H., Kawano, K. & Harrington, T. L. *Vision Res.* **25**, 453–463 (1985).
6. Hyvärinen, J. *The Parietal Cortex of Monkey and Man. Studies of Brain Function* (Springer, Berlin, 1982).
7. Zipser, D. & Andersen, R. A. *Nature* **331**, 679–684 (1988).
8. Duhamel, J., Colby, C. L. & Goldberg, M. E. *Science* **255**, 90–92 (1992).
9. Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. *J. Neurophysiol.* **66**, 1095–1108 (1991).
10. Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W. & Fogassi, L. *J. Neurosci.* **10**(4), 1176–1196 (1990).
11. Andersen, R. A. & Mountcastle, V. B. *J. Neurosci.* **3**, 532–548 (1983).

12. Andersen, R. A., Essick, G. K. & Siegel, R. M. *Science* **230**, 456–458 (1985).
13. Goodman, S. & Andersen, R. A. in *Proc. Int. Joint Conf. Neural Networks, San Diego, Part II*, 381–386 (IEEE Neural Networks Council, Piscataway, New Jersey, 1990).
14. Fogassi, L. et al. *Expl Brain Res.* **89**, 686–690 (1992).
15. Graziano, M. S. A., Yap, G. S. & Gross, C. G. *Science* **266**, 1054–1057 (1994).
16. Andersen, R. A., Essick, G. K. & Siegel, R. M. *Expl Brain Res.* **67**, 316–322 (1987).
17. Gnadt, J. W. & Andersen, R. A. *Expl Brain Res.* **70**, 216–220 (1988).
18. Fetz, E. E. *Behav. Brain Sci.* **15**, 679–690 (1992).
19. Lackner, J. R. & Taublieb, A. B. *Expl Neurol.* **47**, 177–190 (1984).
20. Biguer, B., Donaldson, I. M. L., Hein, A. & Jeannerod, M. *Brain* **111**, 1405–1424 (1988).
21. Taylor, J. L. & McCloskey, D. I. *Brain* **114**, 755–759 (1991).
22. Roll, R., Velay, J. L. & Roll, J. P. *Expl Brain Res.* **85**, 423–431 (1991).
23. Guitton, D. & Volle, M. *J. Neurophysiol.* **58**(3), 427–459 (1987).
24. Fuller, J. H., Maldonado, H. & Schlag, J. *Brain Res.* **271**, 241–250 (1983).
25. Tomlinson, R. D. *J. Neurophysiol.* **64**(6), 1873–1891 (1990).
26. Laurutis, V. P. & Robinson, D. A. *J. Physiol., Lond.* **373**, 209–233 (1986).
27. Snyder, L. H., Brothie, P. R. & Andersen, R. A. *Soc. Neurosci. Abstr.* **19**, 770 (1993).
28. Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. *J. Neurophysiol.* **66**, 1109–1124 (1991).

ACKNOWLEDGEMENTS. We dedicate this paper to the memory of S. J. Goodman, who died on 5 November 1993. We thank D. Bradley for assistance with the statistical analysis; N. Qian and S. Barash for scientific discussions; C. Andersen, L. Rodriguez, and C. Cooper for editorial assistance; and G. Robertson and S. Marchetti for technical assistance. This work was supported by the National Eye Institute and the Office of Naval Research.

Neuronal deficits, not involving motor neurons, in mice lacking BDNF and/or NT4

J. C. Conover*, J. T. Erickson†, D. M. Katz†, L. M. Bianchi*, W. T. Poueymirou*, J. McClain*, L. Pan*, M. Helgren*, N. Y. Ip*, P. Boland*, B. Friedman*, S. Wiegand*, R. Vejsada‡, A. C. Kato‡, T. M. DeChiara* & G. D. Yancopoulos*

* Regeneron Pharmaceuticals, Inc., Tarrytown, New York 10591, USA

† Department of Neurosciences, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA

‡ Department of Pharmacology & Division of Clinical Neuromuscular Research, University Medical Centre, Geneva, Switzerland

NERVE growth factor and other neurotrophins signal to neurons through the Trk family of receptor tyrosine kinases^{1–6}. TrkB is relatively promiscuous *in vitro*, acting as a receptor for brain-derived neurotrophic factor (BDNF), neurotrophin-4 (NT4) and, to a lesser extent, NT3 (refs 3–5). Mice lacking TrkB⁷ show a more severe phenotype than mice lacking BDNF^{8,9}, suggesting that TrkB may act as a receptor for additional ligands *in vivo*. To explore this possibility, we generated mice lacking NT4 or BDNF as well as mice lacking both neurotrophins. Unlike mice lacking other Trks^{7,10,11} or neurotrophins^{8,9,12–14}, NT4-deficient mice are long-lived and show no obvious neurological defects. Analysis of mutant phenotypes revealed distinct neuronal populations with different neurotrophin requirements. Thus vestibular and trigeminal sensory neurons require BDNF but not NT4, whereas nodose-petrossal sensory neurons require both BDNF and NT4. Motor neurons, whose numbers are drastically reduced in mice lacking TrkB, are not affected even in mice lacking both BDNF and NT4. These results suggest that another ligand, perhaps NT3, does indeed act on TrkB *in vivo*.

Homologous recombination targeting vectors in which the BDNF (Fig. 1a) and NT4 coding regions (Fig. 1b) were disrupted and partially deleted were used to generate mice with these mutations at their endogenous BDNF and NT4 loci, respectively (Fig. 1c, d). In agreement with previous studies^{8,9}, mice homozygous for the BDNF mutation (*bdnf*^{−/−}) failed to thrive beyond postnatal day 8 (Fig. 2a), showed an impaired ability to right themselves (Fig. 2c), lacked proper coordination of movement and balance, alternated through periods of hyperactivity and immobility, and generally died by 3 weeks of age. In contrast, mice homozygous for the NT4 mutation (*nt4*^{−/−})

(Fig. 1d), which clearly lacked NT4 expression (Fig. 1e), appeared normal. Growth and righting responses of these mice were similar to those of controls (Fig. 2b, c), and the mice were able to mate and produce viable offspring; these mice have now survived for 10 months without obvious neurological or phenotypic abnormalities. Mice mutant for both BDNF and NT4 (*bdnf*^{−/−} *nt4*^{−/−}) did not appear to be more seriously affected than mice lacking only BDNF; they died within 3 weeks and displayed similar behavioural and growth disturbances. Thus the discrepancy between the phenotypes of mice lacking TrkB (*trkb*^{−/−}), which are severely affected and die within 24 to 48 hours⁷, and mice lacking only BDNF^{8,9} cannot simply be explained by the actions of NT4 on TrkB *in vivo*.

As in *trkb*^{−/−} mice, gross examination of the brains of *bdnf*^{−/−} or *nt4*^{−/−} mice revealed no major cytoarchitectural abnormalities (Fig. 3a), although the brains of *bdnf*^{−/−} mice were smaller than controls and had less neuropile. More detailed analysis may reveal subtle changes in the brains of *nt4*^{−/−} mice similar to the maturational delays and the changes in neuropeptides and calcium-binding proteins observed in *bdnf*^{−/−} mice⁹, as well as behavioural and functional differences.

Examination of various sensory ganglia and motor neuron populations revealed distinct and overlapping requirements for BDNF and NT4. Consistent with the balance and movement abnormalities displayed by *bdnf*^{−/−} mice, but not by *nt4*^{−/−} mice (Fig. 2c), vestibular ganglia were reduced by more than 90% in volume in *bdnf*^{−/−} mice, as previously noted^{8,9}, but appeared normal in *nt4*^{−/−} mice (Fig. 3b and Table 1). The cochlear ganglia, however, were not reduced in size in mice lacking either BDNF or NT4. As with the vestibular ganglia, significant volume reductions (~40%) were seen in the trigeminal ganglia of *bdnf*^{−/−} mice, but not of *nt4*^{−/−} mice; reductions of ganglion size in mice lacking both BDNF and NT4 were not greater than those seen in mice lacking only BDNF (Table 1). However, the situation was quite different for the nodose-petrossal complex, which relays visceral sensory information critical for the regulation of respiration, heart rate, blood pressure and other autonomic functions. The nodose-petrossal complex as a whole was greatly reduced in both *bdnf*^{−/−} and *nt4*^{−/−} mice (~68% and ~61% reductions, respectively, in the volumes of the complex, with similar reductions in the number of neurons in the complex; Table 1 and Fig. 3c). Moreover, an even greater reduction (~83% reduction in volume, and a 79% reduction in neuronal number) was seen in mice lacking both BDNF and NT4 (Table 1 and Fig. 3c).

Previous findings that *trkb*^{−/−} mice had marked reductions in motor neurons of the facial nucleus (70% reduction) and lumbar spinal cord (35% reduction)⁷, but *bdnf*^{−/−} mice did not^{8,9}, led

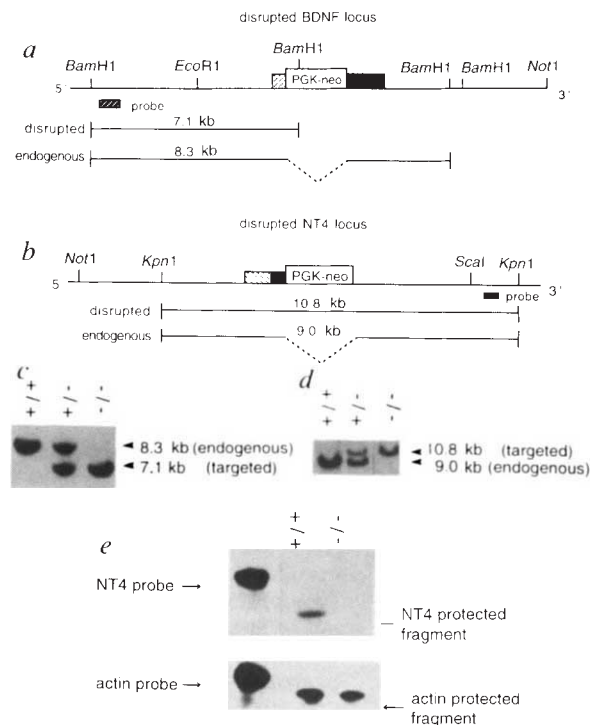
TABLE 1 Volume and cell number of neuron populations

	<i>bdnf</i> ^{+/+} <i>nt4</i> ^{+/+}	<i>bdnf</i> ^{-/-} <i>nt4</i> ^{+/+}	<i>bdnf</i> ^{+/+} <i>nt4</i> ^{-/-}	<i>bdnf</i> ^{-/-} <i>nt4</i> ^{-/-}
	Volume (% reduction)			
Sensory neurons				
Vestibular	24.25±1.86 (n=7)	1.01±0.39 (n=3) (↓96%)	27.31±1.99 (n=3)	Not determined
Trigeminal	284.96±12.67 (n=7)	177.67±11.12 (n=3) (↓40%)	276.80±25.70 (n=5)	186.47±4.21 (n=3) (↓34%)
Nodose-petrosal	38.16±3.82 (n=7)	12.32±2.37 (n=3) (↓68%)	14.93±2.28 (n=5) (↓61%)	6.40±0.38 (n=3) (↓83%)
	Cell number (% reduction)			
Sensory neurons				
Nodose-petrosal	5087±320 (n=7)	1970±238 (n=3) (↓61%)	2240±381 (n=5) (↓56%)	1076±20 (n=3) (↓79%)
	Cell number (% reduction)			
Motor neurons				
Facial motor nucleus	2577±122 (n=8)	2694±158 (n=5)	2719±187 (n=6)	2441±93 (n=3)
Lumbar cord	1923±161 (n=6)	2003±255 (n=3)	1658±247 (n=4)	2059±57 (n=4)

Volumes ($\times 10^6 \mu\text{m}^3$, mean $\pm$ s.e.m.) were calculated, excluding regions occupied by nerve fibre bundles, of different sensory ganglia, and cell counts (mean $\pm$ s.e.m.) were made of the nodose-petrosal complex and the motor neurons of the facial motor nucleus and spinal cord. Most tissues were taken from P14-15 mice, with the exception of 2 nodose-petrosal ganglia from one P16 (*bdnf*^{-/-}) and one P21 (*bdnf*^{-/-} *nt4*^{-/-}) mouse. Tissue sections were prepared as described in Fig. 3 (the trigeminal ganglia were sectioned at a thickness of 7 μm). Each section was examined to identify all sections which contained neuronal somata. These sections were parcellated, from the beginning to the end of the ganglia, into 12 groups (vestibular, trigeminal) and one section from each group was chosen by random sampling²⁹ for further analysis. For the nodose-petrosal ganglia, every second or third section (mutant or wild-type ganglia, respectively) containing neuronal somata was analysed. Morphometric analysis was performed on digitized images of each section using a computer-assisted planimeter to measure the area occupied by neuronal cell bodies (Jandel Image Analysis Systems or NIH Image Analysis Software). The volume of each ganglia occupied by neurons was calculated from the cross-sectional area, section thickness and the number of sections in the ganglion. In addition, corrected counts of nodose-petrosal neurons were obtained. Briefly, all neurons with a nucleus in the plane of section were counted in approximately 20–25% of the sections; a correction for double counts, based on nuclear diameter and section thickness, was applied using Abercrombie's formula³⁰. Facial and lumbar spinal cord motor neurons were counted in serial section (30 μm each); every motor neuron in every section was counted, and no correction factor was applied.

FIG. 1 Targeted disruption of the BDNF and NT4 genes in mouse ES cells and the subsequent generation of *bdnf*^{-/-} and *nt4*^{-/-} mice. Schematized depictions of BDNF (a) and NT4 (b) loci disrupted by insertion of a neomycin-resistant expression cassette (PGK-neo), replacing 332 and 600 base pairs (bp) of the BDNF and NT4 genes, respectively. The protein coding region for both BDNF and NT4 are found on a single exon, which are represented here by the preprocoding region (stippled box) and the mature coding region (solid box). c, d, Genotyping of mice from a *bdnf*^{+/+} or *nt4*^{+/+} cross, respectively. Tail DNA was cut with BamHI (c) or KpnI (d) and hybridized to a probe that recognizes external flanking DNA (indicated by box (a, b)). e, RNase protection assay verifying the absence of NT4 messenger RNA expression in *nt4*^{-/-} mice. The protected fragment (218 bp) is detected in newborn testis RNA from *nt4*^{+/+} mice, but not in newborn testis messenger RNA from *nt4*^{-/-} RNA; testis was chosen because it is the site of highest NT4 RNA expression²⁴.

METHODS. To construct the BDNF replacement vector, a 9.4 kb *Xho*/ *Xba*I mouse genomic DNA fragment (Balb/c) containing the BDNF gene was cloned into a plasmid vector (pBluescript, Stratagene). An *Eco*RI/ *Bgl*II PGK-neo fragment, driven by a phosphoglycerate kinase promoter and flanked with a 3' polyadenylation signal sequence, was modified with *Cla*I linkers allowing compatible-end cloning into the *Nar*I sites of the BDNF coding region, deleting about 300 bp of coding region; for negative selection the thymidine kinase gene (MC1-tk) was ligated upstream to the 5' non-coding region. For the NT4 targeting vector, a plasmid (pBluescript) containing the *Eco*RI/ *Hind*III fragment of PGK-neo was constructed. A 7.0 kb *Not*I/ *Sca*I mouse genomic 5' NT4 fragment (129/J, Clontech) in which the 3' *Sca*I site (located within the coding region of NT4) was modified to an *Eag*I site was cloned into the *Not*I site in the polylinker upstream to the neo gene, while a 6.0 kb *Sph*I/ *Sca*I 3' NT4 fragment (with the *Sph*I site located in the 3' untranslated flank of NT4) was modified with *Cla*I linkers and cloned into the *Cla*I site in the polylinker downstream of the neo gene; the *Xho*I/ *Sal*I fragment of MC1-tk was cloned into the *Xho*I site of the polylinker. 129/ Ola ES cells (E14.1 cells) were electroporated with the linearized (*Not*I) replacement vectors and subjected to selection with G418 (Gibco) and Gancilovir (Ganc, Syntex). Six clones from a total of 68 G418/Ganc resistant clones selected contained the targeted disruption of the BDNF gene, and 7 of 144 G418/Ganc resistant clones selected contained the targeted disruption of the NT4 gene. All 6 ES cell clones disrupted for the BDNF gene were microinjected into C57/Bl6 mouse embryos



and found to contribute to the germ line of chimaeric mice, and all 4 microinjected ES clones disrupted for NT4 gene contributed to the germ line. The resultant F₁ mice heterozygous for each mutation were mated to generate F₂ mice homozygous for these mutations. RNase protection assays (RPAll Ribonuclease Protection Assay Kit, Ambion) were performed on 10 μg total RNA from P1 mouse testis from *nt4*^{+/+} (+/+) and *nt4*^{-/-} (-/-) mice. A 301 bp NT4 probe was synthesized by *in vitro* transcription from a fragment coding 218 bp mouse NT4. Of this 218 bp, 143 bp were from the region deleted in the targeting vector. Actin mRNA levels are shown to indicate loading amounts.

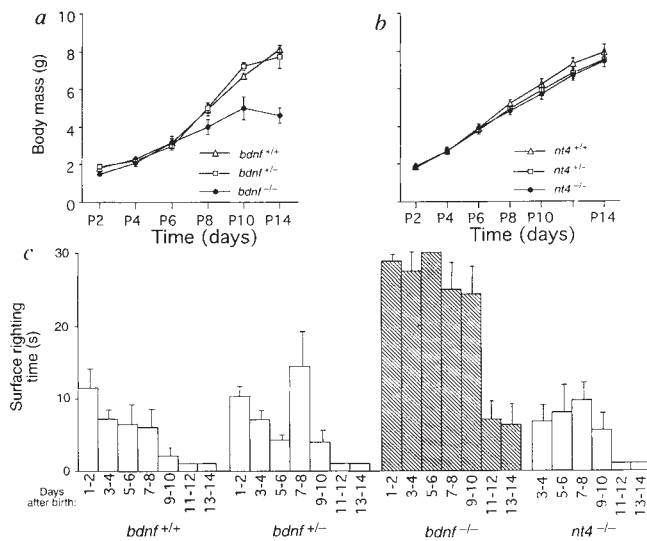


FIG. 2 Growth curves for *bdnf*^{-/-} (a) and *nt4*^{-/-} (b) mice and the development of the righting reflex in *bdnf*^{-/-} and *nt4*^{-/-} mice (c). Bars represent time to complete surface righting. The development of the righting response in *bdnf*^{-/-} mice (stippled bars) is severely impaired compared with *bdnf*^{+/-}, *bdnf*^{+/-} or *nt4*^{-/-} mice.

METHODS. Body mass was taken at approximately the same time each day. To test surface righting reflexes, mice were placed in the supine position and the time for the animal to right itself was recorded. Each mouse was tested twice and the fastest righting time was used for analysis. The criteria for complete righting was that the mouse maintained an upright position for a minimum of 5 s, and 30 s was used as the upper limit for mice unable to right.

to the suggestion that NT4 might be the physiologically relevant TrkB ligand for these cells^{8,9,15}; exogenously provided NT4 is indeed similar to BDNF in its ability to support motor neurons *in vitro* and *in vivo*¹⁶⁻²². However, we found no reduction in motor neurons of the facial nucleus or lumbar cord in *bdnf*^{-/-} or *nt4*^{-/-} mice, or even in *bdnf*^{-/-}*nt4*^{-/-} mice (Fig. 3d and Table 1).

Although previous studies have found that exogenously provided BDNF and NT4 have similar actions on a variety of TrkB-expressing neurons¹⁶⁻²⁵, our analysis has defined distinct *in vivo* roles for these two TrkB ligands. Most of the sensory neurons of the vestibular ganglion, as well as a significant subset of trigeminal sensory neurons, appear to be exclusively dependent upon BDNF; they do not seem to require NT4, nor does NT4

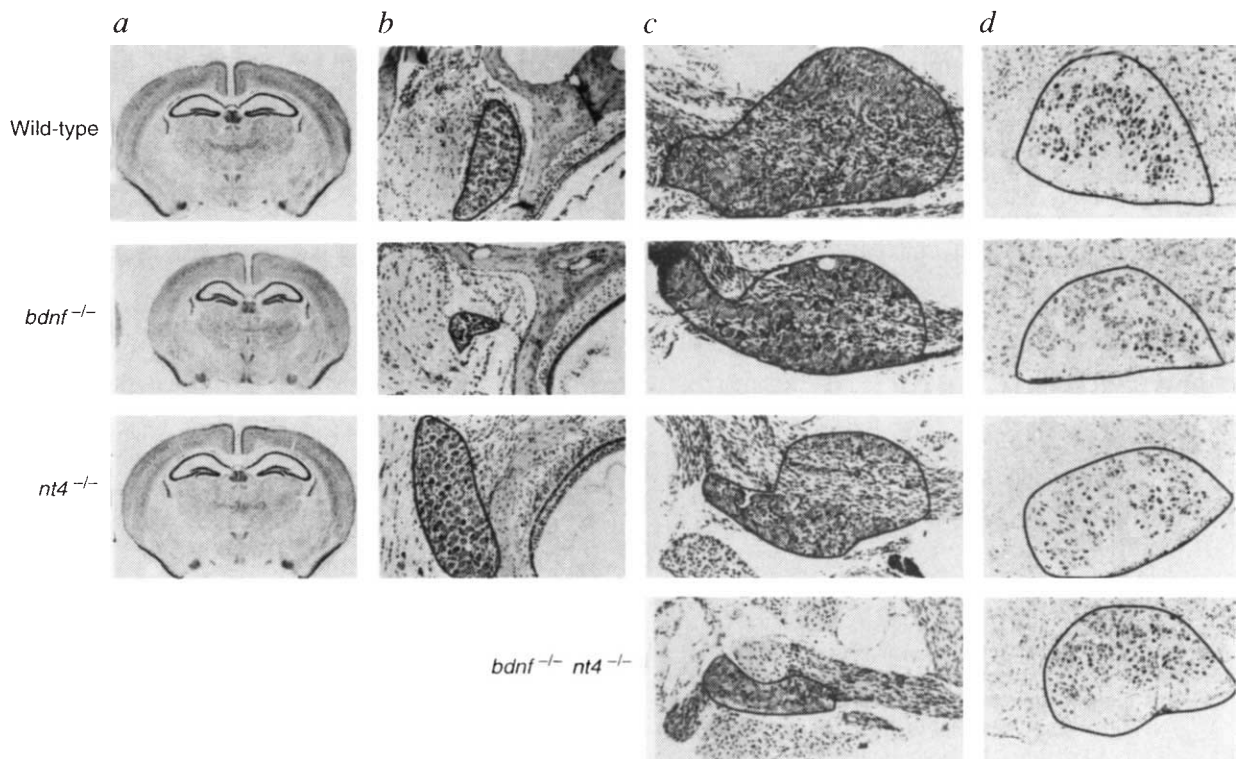


FIG. 3 Morphological analyses of wild-type, *bdnf*^{-/-}, *nt4*^{-/-} and *bdnf*^{-/-}*nt4*^{-/-} mice. **a**, Nissl-stained coronal sections through P15 brains show that the brains of *bdnf*^{-/-} mice, but not *nt4*^{-/-} mice, are clearly smaller than controls. **b**, Comparison of P15 vestibular ganglia shows almost complete lack of vestibular neurons in the *bdnf*^{-/-} mice, but not in the *nt4*^{-/-} mice. **c**, Comparison of the nodose-petrosal ganglia shows a dramatic reduction in the size of the complex in *bdnf*^{-/-} mice, *nt4*^{-/-} mice and a further reduction in the *bdnf*^{-/-}*nt4*^{-/-} mice. **d**, The facial motor nucleus is not decreased in size and appears structurally normal in the *bdnf*^{-/-}, *nt4*^{-/-} and *bdnf*^{-/-}*nt4*^{-/-} mice. Ganglia and nuclei are outlined.

METHODS. Mice were perfused with 4% paraformaldehyde. The inner ear and the region of the carotid bifurcation, including the nodose-petrosal complex and associated nerves (vagus and glossopharyngeal), were embedded in paraffin; the inner ears were decalcified before embedding. The inner ear and nodose-petrosal complex were cut to 8–10 μ m and 7 μ m sections, respectively, and the sections were stained with haematoxylin/eosin. For analysis of the brain and facial motor nucleus, mice were perfused as above, postfixed, and cryoprotected in phosphate buffer with 30% sucrose, and frozen sections were cut to 30 μ m. Brain sections were stained with thionin and sections of facial motor nucleus were stained with cresyl violet.

compensate for the absence of BDNF. However, the nodose-petrossal complex as a whole requires both BDNF and NT4, acting in a largely non-compensatory and non-redundant fashion; large-scale compensation or redundancy between these factors is unlikely because reduction of the complex in mice lacking both BDNF and NT4 was not more severe than the sum of the reductions seen in mice lacking either of these neurotrophins individually. The fact that BDNF and NT4 are differentially expressed in different viscera²⁶ supports the possibility that the subset of visceral afferent innervation remaining intact in *nt4*^{-/-} mice may be quite different from that remaining in *bdnf*^{-/-} mice. The importance of some nodose and petrossal neurons in cardiopulmonary homeostasis has led to the suggestion that loss of these neurons is involved in the early death of *trkb*^{-/-} and *bdnf*^{-/-} mice^{7,9,25}. Although it is perhaps surprising that *bdnf*^{-/-}*nt4*^{-/-} mice do not succumb more quickly than *bdnf*^{-/-} mice, it may be that the subset of visceral afferents supported by NT4 innervate different targets that are not as critical for survival as those dependent upon BDNF.

Motor neurons display a pattern of neurotrophin requirement unlike that of any of the sensory populations examined. Although motor neurons were reported to be drastically reduced in mice lacking TrkB⁷, they are not reduced in *bdnf*^{-/-}*nt4*^{-/-} mice. This result, together with the reported early death of TrkB mutant mice⁷ as compared with mice lacking both BDNF and NT4, suggests that yet another ligand acts via TrkB *in vivo*; however, it must also be considered that the *trkb*^{-/-} and *bdnf*^{-/-}*nt4*^{-/-} mice were generated on different genetic backgrounds and in different laboratory environments. NT3 acting on TrkB may explain differences between *trkb*^{-/-} and *bdnf*^{-/-}*nt4*^{-/-} mice: high concentrations of NT3 can activate TrkB *in vitro*⁵, and it has already been suggested that NT3 is present at sufficient levels during normal development to act via TrkB^{5,27,28}. Because skeletal motor neurons are not reduced in number in NT3 mutant mice^{13,14} or in *bdnf*^{-/-}*nt4*^{-/-} mice, current data could only be explained if these three factors all act on motor neurons via TrkB in a compensatory or redundant fashion, with none of these factors normally limiting; if this were

the case, the loss of motor neurons that normally occurs during embryonic development would then be due to limiting amounts of other target-derived factors.

Finally, the relatively mild deficits thus far associated with lack of NT4 during normal development and early life suggest that NT4 may have a potential role in injury responses or in long-term maintenance of the nervous system. □

Received 16 January; accepted 21 March 1995.

1. Levi-Montalcini, R. *Science* **237**, 1154–1162 (1987).
2. Eide, F. F., Lowenstein, D. H. & Reichardt, L. F. *Expl. Neurol.* **121**, 200–214 (1993).
3. Glass, D. J. & Yancopoulos, G. D. *Trends Cell Biol.* **3**, 262–268 (1993).
4. Barbacid, M. *J. Neurobiol.* **25**, 1386–1403 (1994).
5. Ip, N. Y. *et al. Neuron* **10**, 137–149 (1993).
6. Davies, A. M. *J. Neurobiol.* **25**, 1334–1348 (1994).
7. Klein, R. *et al. Cell* **75**, 113–122 (1993).
8. Ernfors, P., Lee, K.-F. & Jaenisch, R. *Nature* **368**, 147–150 (1994).
9. Jones, K. R., Fariñas, I., Backus, C. & Reichardt, L. F. *Cell* **76**, 989–1000 (1994).
10. Smeyne, R. J. *et al. Nature* **368**, 246–249 (1994).
11. Klein, R. *et al. Nature* **368**, 249–251 (1994).
12. Crowley, C. *et al. Cell* **76**, 1001–1011 (1994).
13. Ernfors, P., Lee, K.-F., Kucera, J. & Jaenisch, R. *Cell* **77**, 503–512 (1994).
14. Fariñas, I., Jones, K. R., Backus, C., Wang, X.-Y. & Reichardt, L. F. *Nature* **369**, 658–661 (1994).
15. Snider, W. D. *Cell* **77**, 627–638 (1994).
16. Yan, Q., Elliott, J. & Snider, W. D. *Nature* **360**, 753–755 (1992).
17. Oppenheim, R. W., Qin-Wei, Y., Prevette, D. & Yan, Q. *Nature* **360**, 755–757 (1992).
18. Sendtner, M., Holtmann, B., Kolbeck, R., Thoenen, H. & Barde, Y.-A. *Nature* **360**, 757–758 (1992).
19. Henderson, C. E. *et al. Nature* **363**, 266–270 (1993).
20. Koliatsos, V. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 3304–3308 (1994).
21. Li, L., Oppenheim, R. W., Lei, M. & Huenou, L. J. *J. Neurobiol.* **15**, 759–766 (1994).
22. Vejsada, R., Sagot, Y. & Kato, A. C. *Eur. J. Neurosci.* **7**, 108–115 (1995).
23. Ibanez, C. F. *et al. Development* **117**, 1345–1353 (1993).
24. Thaler, C. D., Suhr, L., Ip, N. Y. & Katz, D. M. *Dev. Biol.* **161**, 338–344 (1994).
25. Hertzberg, T., Fan, G., Finley, J. C., Erickson, J. T. & Katz, D. M. *Dev. Biol.* **166**, 801–811 (1994).
26. Timmusk, T., Belluardo, N., Metsis, M. & Persson, H. *Eur. J. Neurosci.* **5**, 605–613 (1993).
27. Maisonpierre, P. C. *et al. Neuron* **5**, 501–509 (1990).
28. Squinto, S. P. *et al. Cell* **65**, 885–893 (1991).
29. Pover, C. M. & Coggeshall, R. E. *Anat. Rec.* **231**, 573–578 (1991).
30. Abercrombie, E. M. *Anat. Rec.* **94**, 239–247 (1946).

ACKNOWLEDGEMENTS. We thank L. S. Schleifer, R. Lindsay and others at Regeneron for their comments; F. Alt for the PGK-neo plasmid; K. Thomas for the MC1-tk plasmid; R. Murray for the 129/Ola embryonic stem cells (E14.1); F. Alt, E. Robertson and J. Charron for introducing us to embryonic stem cell culture; C. Ibanez for help with the RNase protection assay; M. Gisser and J. Griffiths for DNA sequencing assistance; E. Buck and C. Kunow for assistance with morphometric analyses; D. Heindl, M. Simmons and P. Guidry for expert technical assistance; and C. Murphy and E. Hubel for graphic and photographic assistance.

Sensory but not motor neuron deficits in mice lacking NT4 and BDNF

Xin Liu, Patrik Ernfors*, Hong Wu & Rudolf Jaenisch

The Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

NEUROTROPHINS play important roles in neuronal survival during vertebrate development^{1–3}. Neurotrophin-4 (NT4), alone or in combination with brain-derived neurotrophic factor (BDNF), has been suggested to be necessary for the survival of peripheral sensory neurons and central nervous system (CNS) neurons, including motor neurons^{4–16}. To define the role of NT4 *in vivo*, we generated mice lacking NT4 by gene targeting. NT4-deficient mice were viable but exhibited a loss of sensory neurons in the nodose-petrossal and geniculate ganglia. In contrast, motor neurons of the facial nucleus and sympathetic neurons of the superior cervical ganglion were unaffected, and there was no obvious loss of dopaminergic

neurons in the substantia nigra. In mice lacking both NT4 and BDNF^{14,15}, facial motor neurons remained unaffected, whereas the loss of sensory neurons was more severe than with either mutation alone. Thus NT4 is required during development for the survival of some peripheral sensory neurons but not sympathetic or motor neurons.

Embryonic stem cells were transfected with a targeting vector containing a NT4 gene genomic DNA fragment in which the entire coding sequence is deleted (Fig. 1a). Targeted embryonic stem cell clones were identified by Southern blot analysis and used to derive NT4 mutant mice (Fig. 1b). Homozygous mutant mice were obtained from two independently targeted embryonic stem cell clones and showed no overt phenotype. To examine the role of NT4 in neuronal survival, serial sections of the heads from day 18.5 embryos and dorsal root ganglia from adults were prepared, and neurons in the facial motor nucleus and in peripheral sensory and sympathetic ganglia were counted. Although no difference in the number of facial motor neurons or superior cervical ganglion neurons was seen, homozygous mutant mice displayed a significant loss of sensory neurons. Approximately half of the neurons were lost in the nodose-petrossal and the geniculate ganglia (Table 1). To determine whether the neuronal deficiency correlated with the loss of a particular phenotypic subpopulation of nodose-petrossal neurons, antibodies against calcitonin gene-related peptide, neuropeptide Y, substance P, tyrosine hydroxylase, somatostatin and the NT4 receptor, TrkB⁸, were used for immunohistochemistry of sections prepared from these ganglia. No obvious differences

* Present address: Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Doktorsringen 12A, BZ-lab, Karolinska Institute, S-171 77 Stockholm, Sweden.

were detected between the mutant and control mice (data not shown).

Because NT4 is a highly potent survival factor for embryonic central dopaminergic neurons in culture¹², and is expressed in the midbrain⁷, neurons in the substantia nigra were examined in adult mutant mice using anti-tyrosine hydroxylase immunohistochemistry. Figure 2 shows that there is no obvious loss of central dopaminergic neurons in NT4 mutant mice (Fig. 2b) as compared with age-matched controls (Fig. 2a). We suggest that NT4 may not be required for the survival of the midbrain dopaminergic neurons, although we cannot exclude subtle changes which would require a rigorous quantitative comparison of mutant and wild-type mice.

Because NT4 and BDNF signal through the same receptor, TrkB^{6,8,17}, the mild phenotype seen in NT4 mutant mice could be explained if BDNF compensated for the lack of NT4. To study whether NT4 and BDNF can functionally substitute for each other, we crossed NT4 with BDNF mutant mice. Mice homozygous for the NT4 mutation and heterozygous for the BDNF mutation were viable and fertile. Most mice homozygous for both mutations, however, died within 48 hours of birth, and only a few survived for up to 10 days. The double-mutant mice were about the same size as their wild-type littermates at birth but failed to grow. They showed defective coordination of movement and balance, which is characteristic for BDNF mutant mice^{14,15}. To examine whether additional neuronal loss occurred in the double-mutant mice, the heads of day 18.5 embryos of NT4 and BDNF double-mutant, BDNF single-mutant, and wild-type embryos were sectioned and neurons counted. In BDNF mutant embryos, the neuronal loss in nodose-petrosal and geniculate ganglia (Table 1) were similar to that previously reported at postnatal day 0 (P0) (ref. 15) or P14 (ref. 14). Interestingly, no statistically significant loss of facial motor neurons was detected in the BDNF-NT4 double mutants (Table 1). However, only $10 \pm 1\%$ and $6 \pm 1\%$ of the neurons were left in the nodose-petrosal and the geniculate ganglia of the double-mutant embryos, respectively, which by far exceeds that seen in the single-mutant mice (Table 1). The number of neurons in other ganglia was also significantly reduced, with $70 \pm 8\%$ remaining in the trigeminal ganglion and $18 \pm 3\%$ in the vestibular ganglion, similar to BDNF mutant mice. We conclude that

lack of both NT4 and BDNF appears not to affect motor-neuron survival but leads to an additive neuronal loss in the nodose-petrosal and geniculate ganglia.

The main defect in mice lacking NT4 was a significant loss of nodose-petrosal and geniculate ganglion neurons. These *in vivo* results are consistent with previous studies showing that NT4 supports the survival of cultured nodose-petrosal neurons⁹. The finding that the loss of sensory neurons in the double mutants is to a large extent additive suggests that the two factors act mostly on distinct and non-overlapping populations of sensory neurons. Because BDNF and NT4 are ligands for the same receptor, our results are consistent with the possibility that different groups of nodose-petrosal or geniculate neurons innervate distinct targets which may selectively express only one or the other factor and thus result in differential survival of neurons in the respective mutant. Another possibility is that different expression of TrkB receptor isoforms^{18,19} in neurons of nodose-petrosal and geniculate ganglia leads to different responses of the neurons towards BDNF or NT4. It is also of interest that mice deficient for NT3 display a neuronal loss of about 50% in nodose-petrosal ganglia^{20,21}. Because virtually all nodose-petrosal neurons are lost in the BDNF-NT4 double-mutant mice, it appears possible that NT3 and BDNF-NT4 may act at different developmental stages to support neuronal survival in nodose-petrosal ganglia, as has been shown previously for other neurons^{22,23}. Alternatively, a simultaneous action of NT3 with BDNF and/or NT4 may be required for the survival of some of the nodose-petrosal neurons.

In contrast to previous studies which showed that NT4 promotes the survival of central dopaminergic and facial motor neurons in culture or in lesion experiments^{12,13}, no defect was found in facial motor nucleus and no obvious loss of dopaminergic neurons was detected in the animals lacking NT4, although our data do not exclude subtle losses in dopaminergic neurons. Our data suggest that *in vitro* results or manipulations in postnatal animals either may not reflect the physiological function of the protein *in vivo*, or that the loss of NT4 is compensated for by other neurotrophic factors. Indeed, the loss of motor neurons in TrkB mutant mice and the lack of a motor-neuron defect in BDNF mutant mice had suggested that NT4 might sustain normal motor-neuron development in BDNF mutant

TABLE 1 Neuronal cell counts in sensory and sympathetic ganglia and facial motor nuclei of wild-type and mutant mice

	Wild-type		<i>nt4</i> ^{-/-}		<i>bdnf</i> ^{-/-}		<i>bdnf</i> ^{-/-} / <i>nt4</i> ^{-/-}	
Ganglion	Mean no. of neurons	% of control	Mean no. of neurons	% of control	Mean no. of neurons	% of control	Mean no. of neurons	% of control
Trigeminal	30,112 ± 1,920 (n = 3)	100 ± 6	29,704 ± 2,312 (n = 3)	98 ± 8			21,168 ± 2,264 (n = 3)	70 ± 8*
Vestibular	2,988 ± 240 (n = 3)	100 ± 8	2,364 ± 152 (n = 3)	79 ± 5	720 ± 96 (n = 4)	24 ± 3**	536 ± 100 (n = 4)	18 ± 3**
Superior cervical	15,564 ± 1,084 (n = 3)	100 ± 7	17,432 ± 1,208 (n = 3)	112 ± 8			14,908 ± 644 (n = 3)	96 ± 4
Nodose-petrosal	4,368 ± 368 (n = 4)	100 ± 8	1,768 ± 96 (n = 4)	41 ± 2**	1,876 ± 220 (n = 4)	43 ± 5**	416 ± 40 (n = 6)	10 ± 1***
Geniculate	912 ± 8 (n = 3)	100 ± 1	452 ± 48 (n = 3)	50 ± 5**	476 ± 108 (n = 4)	52 ± 12**	54 ± 12 (n = 6)	6 ± 1***
Dorsal root (L4)	6,744 ± 272 (n = 3)	100 ± 5	5,828 ± 124 (n = 4)	86 ± 2				
Facial motor neurons	2,788 ± 296 (n = 4)	100 ± 10	2,580 ± 308 (n = 4)	92 ± 11	2,864 ± 312 (n = 4)	102 ± 11	2,480 ± 216 (n = 4)	89 ± 7

Heads from embryos at day 18.5 of gestation were fixed with 4% paraformaldehyde for 24 hours, embedded in paraffin and sectioned at 5 µm thickness. Sections were stained with cresylviolet. Neurons with a clear nucleus and nucleoli were counted in every eighth section. Dorsal root ganglia were dissected from two-month-old mice perfused with 4% paraformaldehyde and postfixed in 4% paraformaldehyde for 16 hours, equilibrated with 30% sucrose and sectioned at 7 µm thickness. Every fourth section was counted. Counts are displayed as number ± s.e.m.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

FIG. 1 Targeted deletion of the NT4 gene (*nt4*) in mouse embryonic stem cells and generation of mutant mice. **a**, Schematic diagram of the strategy used to target *nt4*. Top, restriction map of the *nt4* genomic DNA fragment; boxed area covers the entire coding sequences. Middle, the 11-kb replacement type targeting vector carrying a *neo* gene driven by a phosphoglycerate kinase promoter, which replaced the *Bgl*III–*Hind*III DNA fragment of *nt4*. Bottom, the structure of the targeted allele in which the entire coding sequences of *nt4* were deleted by the homologous recombination; the genomic probe external to the 5' homologous arm is indicated that hybridizes to a 3.5-kb and an 8-kb DNA fragment from mutant and wild-type alleles, respectively. R, *Eco*RV; H, *Hind*III; B, *Bgl*III; S, *Sph*I. **b**, Southern blot of embryonic stem cell clones. DNA from parental J1 embryonic stem cells and independently cloned G418 resistant embryonic stem cell clones (number 25, 26, 27) were digested with *Eco*RV and hybridized to the external probe shown in **a**. Clones containing the expected 3.5-kb *Eco*RV fragment diagnostic for homologous recombinant were obtained at a frequency of 1 in 7. The blot was rehybridized to a *neo* probe to verify a single integration. The blot was also rehybridized to a NT4 cDNA probe to verify further the homologous recombinants in which the intensity of the 8-kb band was reduced to half that from J1 embryonic stem cells. Two independent recombinant embryonic stem cell clones contributed to the germ line of recipient embryos after blastocyst injection. **c**, Southern blot of offspring from a *nt4*^{+/–} × *nt4*^{+/–} cross. Tail DNA was extracted and digested with *Eco*RV and analysed as described in **b**.

METHODS. The J1 embryonic stem cells were obtained and cultured as described²⁹. Following electroporation of embryonic stem cells with 25 µg ml^{–1} of linearized targeting vector at setting of 400 V, 25 µg (Bio-Rad gene pulsar), cells were selected in 0.4 mg ml^{–1} G418. Resident colonies were trypsinized individually. Half of the cells were expanded for freezing, and the other half for genotyping. Mice were generated as described²⁹.

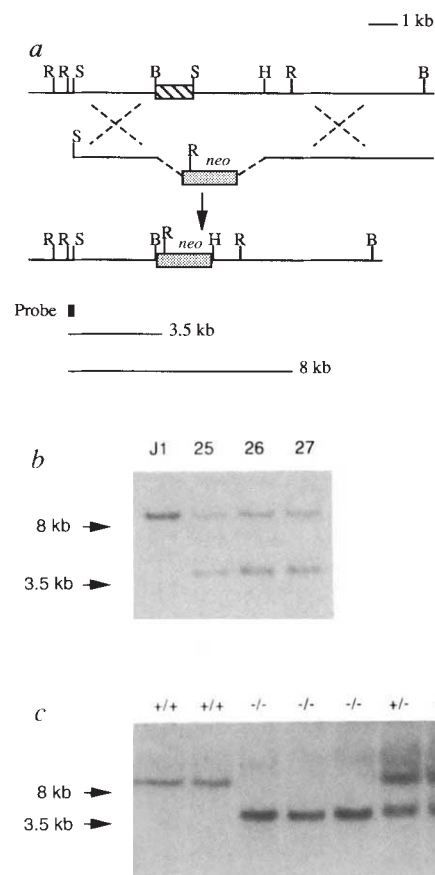
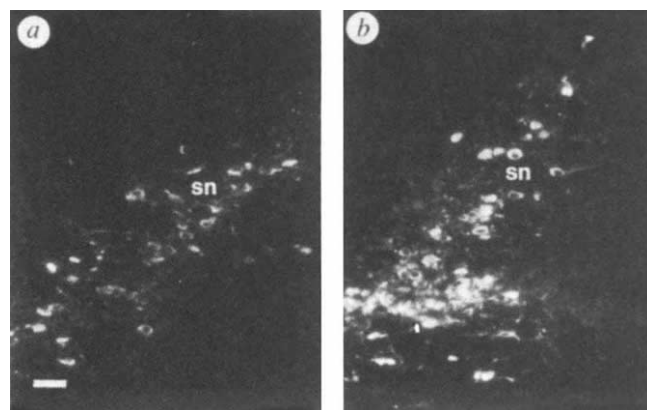


FIG. 2 Tyrosine hydroxylase immunohistochemistry of sections through the midbrain of **a**, wild-type, and **b**, NT4 mutant mice; sn, substantia nigra; scale bar, 10 µm.

METHODS. Four-month-old mice were perfused with 4% paraformaldehyde. Brains were dissected, postfixed for 16 hours, equilibrated with 30% sucrose and sectioned at 7 µm thickness. Sections were preincubated in dilution buffer (0.5 M NaCl, 0.01 M phosphate buffer, pH 7.3, 3% bovine serum albumin and 0.3% Triton X-100) for 1 h followed by overnight incubation with rabbit anti-tyrosine hydroxylase antiserum (Peel-Freeze) diluted at 1:200 in dilution buffer containing 5% goat serum. After 4 washes in PBS, sections were incubated for 2 h with a rhodamine-conjugated goat anti-rabbit secondary antiserum (Cappel) (1:300 dilution).



mice by substituting for BDNF's function^{14,16}. The observation that NT4–BDNF double-mutant mice have no significant loss of facial motor neurons suggests that BDNF and NT4 are not essential for motor-neuron survival, but does not exclude other potential involvement of these two neurotrophins in motor-neuron function. It is also possible that other unidentified factors may activate TrkB. For example, unrelated neurotrophic factors such as GDNF, which has been shown to support motor-neuron survival *in vitro*²⁴, and rescue motor neurons as well as mesencephalic dopaminergic neurons from degeneration *in vivo*^{25–28}, could replace some of the functions of BDNF and NT4 in embryonic development. Because of the potential therapeutic value of neurotrophins in the treatment of neurodegenerative diseases, it will be crucial to elucidate their role in the prenatal and postnatal nervous system. □

Received 1 March; accepted 30 March 1995.

- Levi-Montalcini, R. *Science* **237**, 1154–1162 (1987).
- Barde, Y. A. *Neuron* **2**, 1525–1534 (1989).
- Snider, W. D. *Cell* **77**, 627–638 (1994).
- Berkemeier, L. R. et al. *Neuron* **7**, 857–866 (1991).
- Hallbook, F., Ibanez, C. F. & Persson, H. *Neuron* **6**, 845–858 (1991).
- Ip, N. Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3060–3064 (1992).
- Timmusk, T., Belluardo, N., Metsis, M. & Persson, H. *Eur. J. Neurosci.* **5**, 605–613 (1993).
- Klein, R., Lamballe, F., Bryant, S. & Barbacid, M. *Neuron* **8**, 947–956 (1992).
- Davies, A. M. et al. *J. Neurosci.* **13**, 4961–4967 (1993).
- Ibanez, C. F. et al. *Development* **117**, 1345–1353 (1993).
- Fischer, W., Sirevaag, A., Wiegand, S. J., Lindsay, R. M. & Bjorklund, A. *Proc. natn. Acad. Sci. U.S.A.* **91**, 8607–8611 (1994).
- Hyman, C. et al. *J. Neurosci.* **14**, 335–347 (1994).
- Koliatsos, V. E. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3304–3308 (1994).
- Ernfors, P., Lee, K. F. & Jaenisch, R. *Nature* **368**, 147–150 (1994).
- Jones, K. R., Farinas, I., Backus, C. & Reichardt, L. F. *Cell* **76**, 989–999 (1994).
- Klein, R. et al. *Cell* **75**, 113–122 (1993).
- Ip, N. Y. & Yancopoulos, G. D. *Ann. Neurol.* **35**, S13–16 (1994).
- Klein, R., Conway, D., Parada, L. F. & Barbacid, M. *Cell* **61**, 647–656 (1990).
- Middlemas, D. S., Lindberg, R. A. & Hunter, T. *Molec. cell. Biol.* **11**, 143–153 (1991).

20. Ernfor, P., Lee, K. F., Kucera, J. & Jaenisch, R. *Cell* **77**, 503–512 (1994).
 21. Farinas, I., Jones, K. R., Backus, C., Wang, X. Y. & Reichardt, L. F. *Nature* **369**, 658–661 (1994).
 22. Buchman, V. L. & Davies, A. M. *Development* **118**, 989–1001 (1993).
 23. Di Ciccio-Bloom, E., Friedman, W. J. & Black, I. B. *Neuron* **11**, 1101–1111 (1993).
 24. Henderson, C. E. *et al. Science* **266**, 1062–1064 (1994).
 25. Tomac, A. *et al. Nature* **373**, 335–339 (1995).
 26. Beck, K. D. *et al. Nature* **373**, 339–341 (1995).
 27. Yan, Q., Matheson, C. & Lopez, O. T. *Nature* **373**, 341–344 (1995).
 28. Oppenheim, R. W. *et al. Nature* **373**, 344–347 (1995).
 29. Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915–926 (1992).

ACKNOWLEDGEMENTS. We thank J. Dausman and R. Curry for technical assistance, and S. Dube for help with cell counts. This research was supported by AMGEN (R.J.), the Swedish Medical Research Council, the Swedish Cancer Society, and the Petrus and August Hedlunds Stiftelse (P.E.). X.L. was supported by a fellowship from the Leukemia Society.

Spontaneous resistance to acute T-cell leukaemias in TCRV γ 1.1J γ 4C γ 4 transgenic mice

Josef Martin Penninger*, Tao Wen*, Emma Timms*, Julia Potter*, Valerie A. Wallace*, Toshifumi Matsuyama*, David Ferrick†, Beate Sydora‡, Mitchell Kronenberg‡ & Tak W. Mak*

* Amgen Institute, Ontario Cancer Institute, and Departments of Medical Biophysics and Immunology, University of Toronto, 500 Sherbourne Street, Toronto, Ontario M5X 1K9, Canada

† Department of Veterinarian Microbiology and Immunology, University of California at Davis, Davis, California 95616, USA

‡ Department of Microbiology and Immunology, UCLA Centre for the Health Sciences, Los Angeles, California 90024, USA

THE concept of tumour surveillance implies that specific and non-specific components of the immune system eliminate tumours in the early phase of malignancy^{1,2}. The immunological mechanisms that control growth of preneoplastic cells are, however, not known. T cells expressing $\gamma\delta$ T-cell receptors (TCR) were first described as lymphocytes with reactivity against various tumour cells, which suggests that $\gamma\delta$ T cells could mediate tumour surveillance^{3–6}. Here we show that TCRV γ 1.1J γ 4C γ 4 transgenic mice⁷ are spontaneously resistant to acute T-cell leukaemias but cannot reject non-haematopoietic tumours. TCRV γ 1.1J γ 4C γ 4⁺ hybridomas isolated from these mice react *in vitro* against almost all haematopoietic tumour cell lines tested. Recognition of tumour cells depends on the $\gamma\delta$ TCR but is independent of major histocompatibility complex (MHC) class I, MHC class II, or TAP-2 peptide transporter expression. Ligand recognition is influenced by the murine *Nrmp* gene, which confers resistance or susceptibility to tuberculosis, lepra and leishmaniasis^{8,9}. These data indicate that TCRV γ 1.1⁺ T cells confer spontaneous immunity against haematopoietic tumours *in vivo* and link innate resistance to bacterial infections with tissue-specific tumour surveillance by $\gamma\delta$ ⁺ T cells.

To determine the function of circulating $\gamma\delta$ T cells, we injected TCRV γ 1.1J γ 4C γ 4 transgenic mice⁷ with two T-cell leukaemia cell lines, ALC and RBL-5 (Fig. 1). The TCRV γ 1.1 chain is the most prominent TCR γ chain expressed on T cells in the spleen, lymph nodes and peripheral blood of mice, but is only sparsely found in other tissues^{3–6}. Injection of ALC or RBL-5 leukaemia cells leads to tumour growth in C57Bl/6 syngenic mice, defined by ascites and massive infiltration of the kidney, intestine, liver, spleen and lymph nodes. TCRV γ 1.1 transgenic mice were spontaneously resistant against both tumours as compared with TCRV γ 1.1 non-transgenic littermates and C57Bl/6 mice (Fig. 1). We tested three TCRV γ 1.1 transgenic mouse lines containing different copy numbers and transgenic integration sites. In two mouse strains containing high and intermediate copy numbers

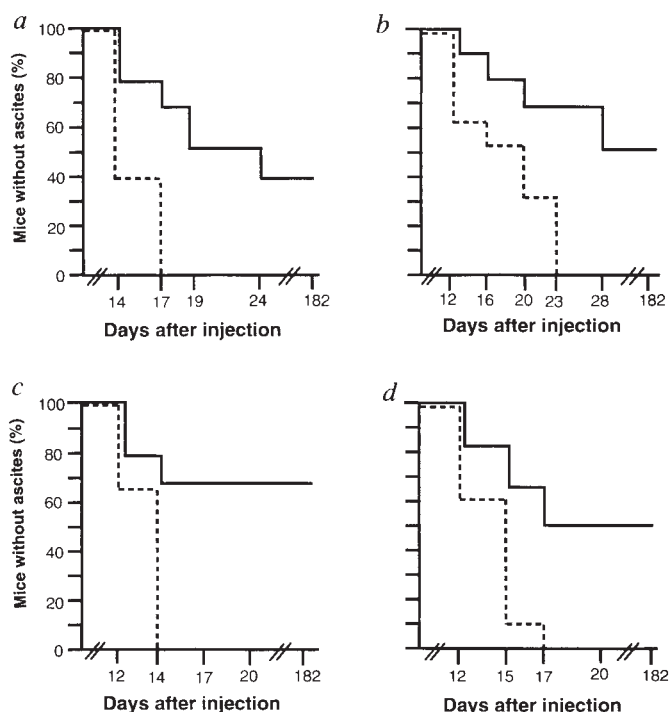


FIG. 1 Spontaneous resistance to RBL-5 (a and b) and ALC (c and d) leukaemia cells in TCRV γ 1.1J γ 4C γ 4 transgenic mice. a, 1000 RBL-5 cells per C57Bl/6 (broken line) or syngenic TCRV γ 1.1J γ 4C γ 4 (solid line) mice; 8 mice per group. b, 500 RBL-5 cells per TCRV γ 1.1J γ 4C γ 4 transgenic-negative (broken line) or TCRV γ 1.1J γ 4C γ 4 transgenic-positive (solid line) littermates; 9 mice per group. c, 1000 ALC cells per C57Bl/6 (broken line) or TCRV γ 1.1J γ 4C γ 4 transgenic (solid line) mice; 9 mice per group. d, 500 ALC cells per C57Bl/6 (broken line) or syngenic TCRV γ 1.1J γ 4C γ 4 Tg (solid line) mice; 10 mice per C57Bl/6 group and 16 mice per TCRV γ 1.1⁺ group. a, c, Very high copy number strain (~200 transgenic copies). b, d, Intermediate copy number strain (~20 transgenic copies). Data from the low copy number strain (~10 copies) are not shown⁷. Mice were followed for 6 months after tumour injection (day 182).

METHODS. Five-week-old TCRV γ 1.1⁺ mice (6th backcross into a C57Bl/6 background) and age-matched C57Bl/6 mice or TCRV γ 1.1⁺ and TCRV γ 1.1⁺ littermates were injected intraperitoneally with ALC or RBL-5 tumour cells, and the development of the tumour (ascites) was followed daily. Spontaneous resistance to tumour cells is dependent on the dose of tumour cells (not shown). Mice with ascites were killed for ethical reasons. To determine whether syngenic TCRV γ 1.1⁺ transgenic mice were immunologically compatible with C57Bl/6 mice, all transgenic mice resistant to RBL-5 or ALC tumours were transplanted with C57Bl/6 skin. We never observed skin rejection. In addition, RBL-5 and ALC tumour cells were not recognized by NK cells (not shown). Expression of the transgene was determined by genomic Southern blotting of EcoRI-digested tail DNA using a TCRC γ 4 probe (not shown). All experiments conformed to Guidelines of the Canadian Research Council, and were approved by the Animal Care Committee of the Ontario Cancer Institute.

of the transgene the V γ 1.1 transgene conferred spontaneous resistance to ALC and RBL-5 (Fig. 1), whereas the strain containing a low copy number was not resistant against acute T leukaemias. Thus anti-tumour activity of TCRV γ 1.1⁺ cells depends on the expression level of the transgene and is not linked to the transgenic integration site.

ALC tumour cells expressing the clonotypic TCRV β 12 chain could be detected in the blood of TCRV γ 1.1 transgenic mice 3 days after inoculation. In parallel, an expansion of $\gamma\delta$ ⁺ T cells was detected in the spleen of TCRV γ 1.1 transgenic and C57Bl/6 mice after injection of leukaemic T cells (not shown). Transgenic mice expressing the TCRV γ 2J γ 1C γ 1 chain did not reject ALC or RBL-5 T cell leukaemias (not shown). In addition, growth kinetics of two C57Bl/6-derived fibrosarcomas (MC57X and

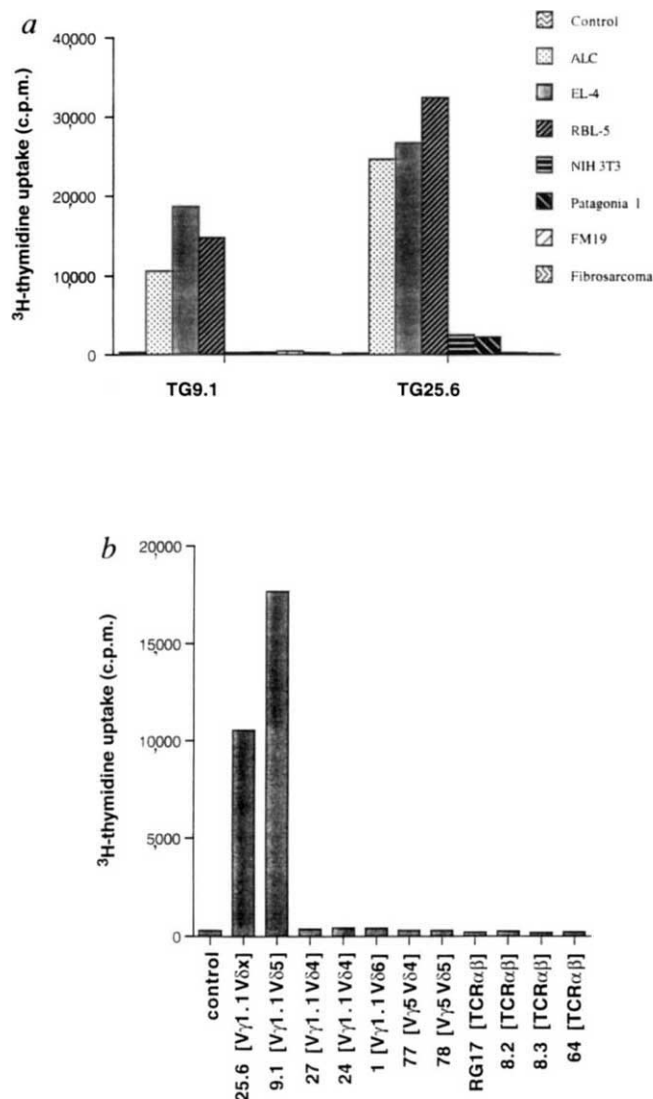


FIG. 2 Reactivity of different TCR $\gamma\delta^+$ hybridomas against a various haematopoietic T-cell leukaemia (ALC, EL-4 and RBL-5) and non-haematopoietic fibroblast lines (NIH-3T3; FM19; Patagonia 1; MC57X fibrosarcoma), and *b* P388D lymphoid tumour cells. For description of stimulator cells see legend to Table 1. Spontaneous IL-2 release of TG9.1 (*a* and *b*) and TG25.6 (*a*) in the absence of stimulator cells are shown as controls. IL-2 production of all other T-cell hybridomas and all irradiated stimulator cells was similar to the indicated controls (not shown). Expression of TCR $\gamma\delta$ and TCR $\alpha\beta$ chains was determined before the stimulation assay, and all hybridomas could be activated (IL-2 production) using monoclonal antibodies against the TCR $\gamma\delta$ (clone GL-3) or the TCR $\alpha\beta$ (clone H57-597.2.1). One representative experiment out of more than 10 is shown.

METHODS. TCR hybridomas were derived from six-week-old TCRV γ 4C γ 4 transgenic mice⁷ through fusion of TCR $\gamma\delta^+$ or TCR $\alpha\beta^+$ T cells with the TCR $\alpha\beta$ -deficient variant of the AKR thymoma line BW5147. Hybridomas were selected in HAT medium and screened for TCR $\alpha\beta$ or TCR $\gamma\delta$ expression using monoclonal antibodies. Expression of different TCRV γ and TCRV δ chains was determined using polymerase chain reaction (PCR). To determine the reactivity of TCR $\gamma\delta^+$ and TCR $\alpha\beta^+$ hybridomas, 5×10^3 responder T cells were incubated with 2×10^4 irradiated (4500 rad) stimulator cells in a final volume of 200 μ l Iscove's modified Dulbecco's medium (IMDM) containing 10% fetal calf serum and 10^{-5} M β -mercaptoethanol. After 48 h, supernatants from triplicate cultures were collected and added to 5×10^3 IL-2-dependent CTLL-2 cells at various dilutions (50%, 25%, 12% and 6%). CTLL-2 cells were then cultured for 48 h and pulsed with 3 H-thymidine (1 μ Ci per well) for 8 h. 3 H-thymidine uptake by CTLL-2 cells was determined using liquid scintillation (Beckman).

MC57T) and of a melanoma (B16F1) were not altered in TCRV γ 1.1 transgenic mice as compared with non-transgenic littermates or C57Bl/6 mice (not shown). These data show that tumour growth correlates with expansion of $\gamma\delta$ T cells, and that TCRV γ 1.1⁺ T cells display spontaneous reactivity against T-cell leukaemias *in vivo*.

To determine the mechanism for spontaneous tumour resistance, we generated T-cell hybridomas expressing the TCRV γ 1.1J γ 4C γ 4 chain. Two hybridomas (TG9.1 and TG25.6) responded *in vitro* to ALC, RBL-5 and EL-4 T-cell leukaemia lines, but did not respond to non-haematopoietic cells (Fig. 2*a*). In addition, these hybridomas responded against most haematopoietic tumour cell lines tested including P388D lymphoid tumours and P815 mastocytoma cells (Fig. 2*b*; Table 1). Reactivity against haematopoietic tumour cells was restricted to hybridomas expressing the TCRV γ 1.1 chain, and could not be observed with TCRV γ 5⁺ or TCR $\alpha\beta$ ⁺ hybridomas (Fig. 2*b*). However, not all TCRV γ 1.1⁺ hybridomas responded against tumour cells *in vitro* (Fig. 2*b*).

Most $\gamma\delta$ T cells recognize ligands independent of MHC restriction^{3,6} and can develop in β 2-microglobulin (β 2-m)-deficient mice that lack MHC class I expression¹⁰. In addition, the molecular contact sites between two TCR $\gamma\delta^+$ cell lines and MHC molecules are different from TCR $\alpha\beta$ /MHC interactions¹¹. To determine the ligand for tumour-reactive TCRV γ 1.1⁺ hybridomas, we tested their reactivity against lipopolysaccharide-activated spleen cells isolated from different mouse strains and β 2-m^{-/-} mice (Table 1). TCRV γ 1.1⁺ hybridomas could be stimulated by target cells independent of the MHC haplotype

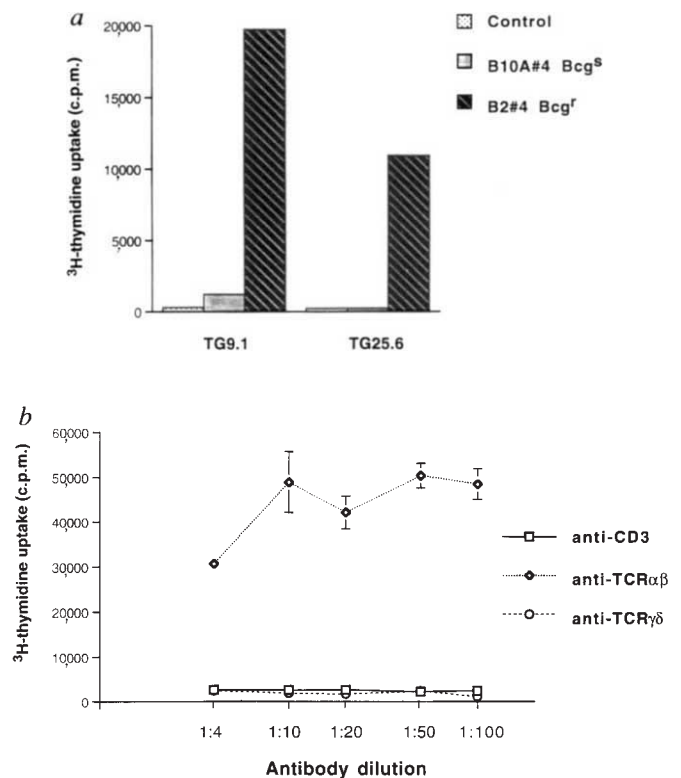


FIG. 3 Reactivity of TCRV γ 1.1⁺ hybridomas against congenic Bcg^S B10A#4 and Bcg^f B2#4 macrophages⁸ (*a*). Controls indicate spontaneous IL-2 release of TG9.1 or TG25.6 in the absence of stimulator cells. *b*, The response of the TCRV γ 1.1⁺ hybridoma TG9.1 against the Bcg^f B2#4 macrophage can be blocked using anti-TCR $\gamma\delta$ (GL-3, hamster IgG) or anti-CD3 (145-2C11, hamster IgG) antibodies, but not with anti-TCR $\alpha\beta$ (H57-597.2.1, hamster IgG). Stimulation and CTLL-2 assays were as in Fig. 2. For blocking studies, stimulator cells were incubated with various concentrations of antibody supernatants before the addition of TG9.1 stimulator cells (at 37 °C for 2 h). One representative experiment is shown.

TABLE 1 Reactivity of TCRV γ 1.1J γ 4C γ 4⁺ T-cell hybridomas

Stimulator cells	Reactivity of TCRV γ 1.1 ⁺ hybridomas		Stimulator cells	Reactivity of TCRV γ 1.1 ⁺ hybridomas	
	Reactivity	SI		Reactivity	SI
Lymphoid tumours			Macrophages		
P388D-H2 ^{d/d}	+++	58	CJA (J774G8- <i>M. leprae</i> HSP65 transfected)	++	46
P388D+rHSP65 (<i>M. tub.</i>)	+++	53	CJK (J774G8- <i>M. leprae</i> HSP65 transfected)	+++	84
P388D+p180-196 <i>M. tub.</i> HSP65 peptide	++	30	EJ (J774G8- <i>M. leprae</i> HSP65 inverse)	+++	55
Bruce Chatwin 3 (P388D- <i>M. tub.</i> HSP65 transfected)	+++	52	B2 no. 4 (Bcg ^f)	+++	136
Bruce Chatwin 4 (P388D- <i>M. tub.</i> HSP65 transfected)	+++	73	B10A no. 4 (Bcg ^s)	+/-	3
Pink14 (P388D- <i>M. tub.</i> HSP65 inverse)	+++	72	TIB186-H2 ^{b/b}	+/-	2
EL4-H2 ^{b/b}	+++	91	Non-haematopoietic cells		
ALC-H2 ^{b/b}	+++	84	NIH 3T3 embryonic fibroblasts	-	1
RBL-5-H2 ^{b/b}	+++	110	NIH 3T3+rHSP65 (<i>M. tub.</i>)	-	1
RMA	++	11	NIH 3T3+p180-196 <i>M. tub.</i> HSP65 peptide	-	1
RMA-S (TAP-2 deficient)	++	18	Patagonia 1 (NIH 3T3- <i>M. tub.</i> HSP65 transfected)	-	1
38B9tk ⁻	+	7	Tierra del Fuego 2 (NIH 3T3- <i>M. tub.</i> HSP65 transfected)	-	1
BW	+/-	5	FM-19 (c-fms-transformed NIH 3T3)	-	1
Mastocytoma			MC57X (fibrosarcoma)	-	1
P815-H2 ^{d/d}	+++	62	MC57T (fibrosarcoma)	-	1
LPS-activated spleen cells			Melanoma	-	1
C3H-H2 ^{k/k}	+	6			
B10.Br-H2 ^{k/k}	+/-	4			
CD1-H2 ^{a/q}	+	9			
BALB/c-H2 ^{d/d}	+	10			
C57Bl/6-H2 ^{b/b}	+	6			
β 2-m ^{-/-} H2 ^{b/b}	+	8			

TCRV γ 1.1⁺ hybridomas were stimulated with various haematopoietic and non-haematopoietic tumour cell lines as described in Fig. 1. Data from more than 15 experiments are compiled. All tumour-reactive TCRV γ 1.1⁺ hybridomas described in Fig. 1 showed a similar pattern of reactivity. Because we have tested only a limited number of non-haematopoietic cell lines, reactivity of TCRV γ 1.1⁺ cells against certain non-haematopoietic cells cannot be excluded completely. The HSP65 peptide (amino acids 180-196) of *M. tuberculosis* (*M. tub.*) HSP65 and the *M. tub.* rHSP65 were used as described^{13,16}. Culture supernatants of stimulator cells had no effect on the response of TCRV γ 1.1⁺ hybridomas. Reactivity (average stimulation index (SI) of different experiments): +++, >50 $\times$ SI; ++, 10-50 $\times$ SI; +, 5-10 $\times$ SI; +/-, 2-5 $\times$ SI. Experimental SI numbers from the hybridoma TG9.1 are indicated on the right. SI is the thymidine uptake of CTLL-2 cells cultured with supernatant from TCRV γ 1.1⁺ hybridomas with stimulator cells/thymidine uptake of CTLL-2 cells cultured with supernatant from TCRV γ 1.1⁺ hybridomas without stimulator cells. Stimulator cells were: (1) P388D: methylcholanthrene-induced lymphoid neoplasma from DBA/2 mice (ATCC TIB 63); (2) Bruce Chatwin 3, Bruce Chatwin 4, and Pink 14: P388D tumours stably transfected with sense or anti-sense *M. tub.* HSP65 complementary DNA using the spleen focus-forming virus (SFFV) 5' LTR as promoter. Expression of the *M. tub.* HSP65 protein was detected in the cytoplasm with various *M. tub.* HSP65-specific monoclonal antibodies (J.M.P., unpublished data); (3) EL4: 9,10-dimethyl-1,2-benzanthracene-induced lymphoma from C57Bl/6 mice (ATCC TIB 39); (4) RBL-5 cells: Rauscher virus-induced leukaemia cells from C57Bl/6 mice. RMA and RMA-S¹² were derived by mutagenesis of RBL-5 cells. RMA-S cells have a profound, albeit incomplete, defect in MHC class I peptide presentation resulting from a mutation in the endoplasmic TAP-2 peptide transporter; (5) ALC: D-RadLV-transformed T-cell leukaemia cells from C57Bl/6 mice that express the TCRV β 12 chain which allows observation of tumour development *in vivo* (J.M.P., unpublished data); (6) 38B9tk-: B-cell lymphoma; (7) BW (BW5147) AKR thymoma cells: see legend to Fig. 1; (8) P815 mastocytoma: derived from DBA/2 mice (ATCC TIB 64); (9) lipopolysaccharide (LPS) (20 μ g ml⁻¹) activated spleen cells were derived from the indicated mouse strains (Jackson Laboratories); (10) CJA, CJK and EJ cells: J774 macrophages (derived from a spontaneous tumour in a Balb/c mouse) transfected with *M. leprae* HSP65 sense or antisense cDNA¹⁵; (11) B10A no. 4 and B2 no. 4 macrophages: derived from Bcg congenic B10.A mice⁸; (12) TIB186 cells (IC-21): transformed peritoneal C57Bl/6 macrophage line; (13) NIH 3T3: contact-inhibited fibroblasts from Swiss mouse embryos (ATCC CRL1658); (14) Patagonia 1 and Tierra del Fuego: NIH 3T3 fibroblasts stably transfected with *M. tub.* HSP65 (J.M.P., unpublished data); (15) FM-19 fibroblasts: c-fms transformed NIH 3T3 cells (V.A.W., unpublished data); (16) MC57X and MC57T: methylcholanthrene-induced fibrosarcomas from C57Bl/6 mice; (17) B16F1: malignant melanoma derived from C57Bl/6 mice.

and β 2-m expression. The *in vitro* reactivity of TCRV γ 1.1⁺ hybridomas was also independent of the endoplasmic TAP-2 peptide transporter, as these cells responded to RMA cells and TAP-2-defective RMA-S cells¹² (Table 1). Reactivity of TCRV γ 1.1⁺ hybridomas was abrogated with anti-CD3 and anti-TCR γ δ monoclonal antibodies but not with those against MHC class I (H-2K, H-2D, TIA^b), MHC class II (I-A, I-E), TCR- α β , or HSP65 proteins (not shown and Fig. 3b). Stable expression of *Mycobacterium tuberculosis* or *M. leprae* HSP65 (refs 13-15) had no influence on the reactivity of TCRV γ 1.1⁺ hybridomas against P388D, J774 or NIH 3T3 cells (Table 1; Fig. 2a). Moreover, addition of a short HSP65 peptide¹⁶ or recombinant *M. tuberculosis* HSP65 to stimulator cells (Table 1), or various heat-shock regimes before stimulation (not shown), did not alter the response. These data imply that recognition of tumour cells through the TCRV γ 1.1 chain is independent of class I and class II MHC, the TAP-2 peptide transporter, and heat-shock proteins.

Because $\gamma\delta$ ⁺ T cells have a role during infection with intracellular bacteria such as *Listeria monocytogenes*^{17,18}, *M.*

tuberculosis^{19,20} or *M. leprae*²¹, we tested the reactivity of TCRV γ 1.1⁺ hybridomas against Bcg congenic macrophages⁸. Bcg is encoded by the *Nromp* gene which has homology to prokaryotic and eukaryotic ABC transporters^{9,22}, and one amino-acid substitution of the *Nromp* protein confers resistance (Bcg^r) or susceptibility (Bcg^s) to infections with *Mycobacteria*, *Salmonella typhimurium* and *Leishmania donovani*^{8,9}. TCRV γ 1.1⁺ hybridomas responded to Bcg^r macrophages but not to the Bcg^s congenic macrophage line (Fig. 3a). Reactivity against Bcg^r macrophages was restricted to TCRV γ 1.1⁺ cells (not shown) and could be blocked with anti-CD3 and anti-TCR γ δ monoclonal antibodies but not with those against TCR α β , MHC class II or MHC class I (Fig. 3b and not shown). In addition, TCRV γ 1.1⁺ hybridomas responded against Bcg^r J774 macrophages but not against the Bcg^s macrophage line TIB186 (Table 1). Because *Nromp* may function as a lysosomal nitrogen transporter^{9,22}, the role of *Nromp* in the recognition by $\gamma\delta$ T cells might be indirect through the generation of nitric oxide involved in signal transduction and/or macrophage effector functions²³. It also cannot be excluded that *Nromp* functions as a peptide transporter from

lysosomes to the cytoplasmic compartment. It should be noted that, although TCRV γ 1.1⁺ T-cell recognition of macrophages correlates with a *Nromp* point mutation, the same TCRV γ 1.1⁺ hybridomas can recognize cells such as P388D tumour cells which do not express the pre-B-cell isoform of *Nromp* (Table 1; Fig. 2b; ref. 9). However, these data provide clear evidence that the *Bcg* locus controls recognition by tumour-reactive $\gamma\delta$ ⁺ cells *in vitro*.

Our data show that TCRV γ 1.1⁺ T cells respond against almost all haematopoietic tumour cells *in vitro* and control growth of two T-cell leukaemias *in vivo*, suggesting that $\gamma\delta$ ⁺ cells mediate tissue-specific tumour immunosurveillance. In addition, we demonstrate that the *Bcg* locus, which confers innate resistance to intracellular infections, can control $\gamma\delta$ recognition *in vitro* and may link tumour-reactive $\gamma\delta$ T cells to $\gamma\delta$ reactivity against intracellular parasites. □

Received 1 December 1994; accepted 24 March 1995.

1. Klein, G. & Boon, T. *Curr. Opin. Immun.* **5**, 687–692 (1993).
2. Pardoll, D. M. *Curr. Opin. Immun.* **5**, 719–725 (1993).
3. Haas, W., Pereira, P. & Tonegawa, S. A. *Rev. Immun.* **11**, 637–686 (1993).

4. Allison, J. P. & Havran, W. L. A. *Rev. Immun.* **9**, 679–705 (1991).
5. Raulet, D. H. A. *Rev. Immun.* **7**, 175–207 (1989).
6. Kronenberg, M. *Curr. Opin. Immun.* **6**, 64–71 (1994).
7. Ferrick, D. A. et al. *Cell* **57**, 483–492 (1989).
8. Radzioch, D. et al. *J. Leuk. Biol.* **50**, 263–270 (1991).
9. Vidal, S. M., Maio, D., Vogan, K., Skamene, E. & Gros, P. *Cell* **73**, 469–485 (1993).
10. Correa, I. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 653–657 (1992).
11. Schild, H. et al. *Cell* **76**, 29–37 (1994).
12. Kärre, K., Ljunggren, H. G., Piöntek, G. & Kiessling, R. *Nature* **319**, 675–678 (1986).
13. O'Brien, R. L. et al. *Cell* **57**, 667–674 (1989).
14. Yeung, R. A. & Elliot, T. J. *Cell* **59**, 5–8 (1989).
15. Silva, C. L., Lukacs, K. & Lowrie, D. B. *Immunology* **78**, 35–42 (1993).
16. Born, W. et al. *Science* **249**, 67–69 (1990).
17. Munk, M. E., Gatrill, A. J. & Kaufmann, S. H. E. *J. Immun.* **145**, 2434–2439 (1990).
18. Mombaerts, P., Arnoldi, J., Russ, F., Tonegawa, S. & Kaufmann, S. H. E. *Nature* **365**, 53–56 (1993).
19. Kabelitz, D. et al. *J. exp. Med.* **173**, 1331–1338 (1991).
20. De Libero, G. et al. *J. exp. Med.* **173**, 1311–1322 (1991).
21. Uyemura, K. et al. *J. exp. Med.* **174**, 683–692 (1992).
22. Barton, C. H., White, J. K., Roach, T. I. A. & Blackwell, J. M. *J. exp. Med.* **179**, 1683–1687 (1994).
23. Nathan, C. *FASEB J.* **6**, 3051–3061 (1992).

ACKNOWLEDGEMENTS. We thank D. Radzioch, R. A. Yeung, D. Lowrie, G. Klein and the WHO for providing reagents, and P. Ohashi, H. W. Mittrücker, R. Schmits, C. Brunaud, J. L. de la Pampa, M. Mora and D. Siderowski for comments. This work was supported by the National Cancer Institute and Medical Research Council of Canada, the NIH (M.K.), the Crohn's and Colitis Foundation of America (B.S.), and the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (J.P.).

Furin-dependent intracellular activation of the human stromelysin-3 zymogen

Duanqing Pei & Stephen J. Weiss*

Division of Hematology/Oncology, Department of Internal Medicine, The University of Michigan Comprehensive Cancer Center, Ann Arbor, Michigan 48109-0640, USA

HUMAN stromelysin-3, a new member of the matrix metalloproteinase family, is expressed in tissues undergoing the active remodelling associated with embryonic development, wound healing and tumour invasion^{1–3}. But like all other members of the matrix metalloproteinase gene family, stromelysin-3 is synthesized as an inactive precursor that must be processed to its mature form in order to express enzymic activity^{4,5}. Here we identify stromelysin-3 as the first matrix metalloproteinase to be discovered that can be processed directly to its enzymically active form by an obligate intracellular proteolytic event that occurs within the constitutive secretory pathway. Intracellular activation is regulated by an unusual 10-amino-acid insert sandwiched between the pro- and catalytic-domains of stromelysin-3, which is encrypted with an Arg-X-Arg-X-Lys-Arg recognition motif for the Golgi-associated proteinase, furin, a mammalian homologue of the yeast Kex2 pheromone convertase^{6,7}. A furin–stromelysin-3 processing axis not only differentiates the regulation of this enzyme from all previously characterized matrix metalloproteinases, but also identifies pro-protein convertases as potential targets for therapeutic intervention in matrix-destructive disease states.

Although previously characterized members of the matrix metalloproteinase (MMP) family are secreted as inactive zymogens^{4,5}, COS-7 cells stably transfected with ST3 complementary DNA (COS 4-2) spontaneously expressed ST3 activity and degraded α_1 -proteinase inhibitor (α_1 PI; a substrate for the active MMP⁸). Proteolysis was completely inhibited by tissue inhibitor of metalloproteinase-1 or -2 (TIMP-1 or TIMP-2, respectively) or the synthetic metalloproteinase inhibitor, BB-94 (ref. 9) (Fig. 1a). Immunoprecipitation with ST3 polyclonal antisera revealed the presence of two major bands in the conditioned medium of ST3-transfected COS cells with approximate

relative molecular masses of 65K and 45K, respectively (Fig. 1b). The 65K species represents the ST3 zymogen, and the 45K species was previously identified as the processed mature form of the enzyme on the basis of an amino terminus homologous to that of the mature form of known MMPs (that is, Phe 98), an intact carboxy terminus, and its ability to cleave the bait region of α_2 -macroglobulin⁸. Importantly, ST3 processing to the active 45K form was similarly observed in ST3-transfected HT-1080 and MCF-7 (Fig. 1b) as well as 293, MDCK and CHO cell lines, where the secreted enzyme was recovered only in its mature form ($n=3$).

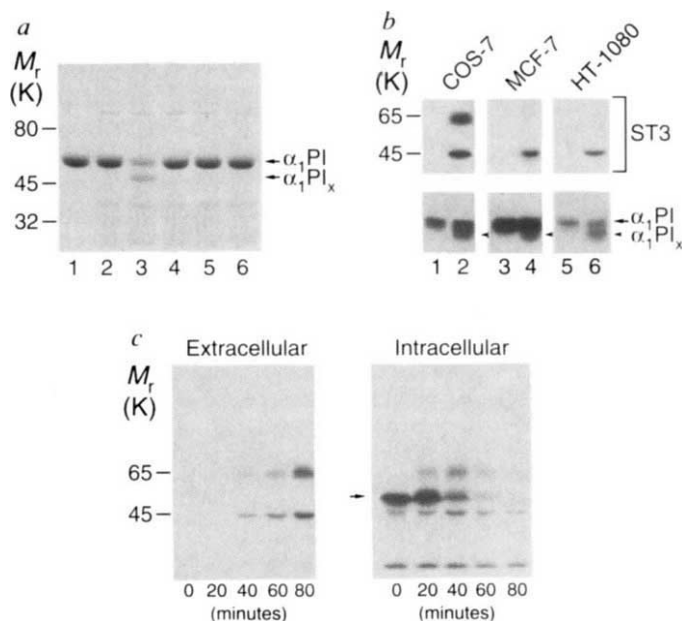
In the extracellular media, pulse-chase analysis of ST3 processing demonstrated that both the 65K zymogen and the 45K active forms of the proteinases were detected in tandem fashion by as early as 40 min post-labelling (Fig. 1c). These results are consistent with a rapid extracellular processing event, but the ratio of active ST3 to zymogen did not increase with continued incubation (Fig. 1c). Furthermore, the activation of the ST3 zymogen was unaffected by proteinase inhibitors capable of preventing processing of all other MMPs^{4,5} (that is, aprotinin, benzamidin, E-64, α_2 -macroglobulin, TIMP-1/2 or BB-94; data not shown). Although the rapid accumulation of active ST3 in the extracellular milieu and the resistance of the processing event to extracellular antiproteinases are consistent with an intracellular activation cascade, MMPs have been assumed only to undergo activation after secretion^{4,5}. However, when pulse-chase analyses of the intracellular pool of ST3 were examined, the 45K processed form of the proteinase could be detected before its secretion into the extracellular medium (Fig. 1c).

The intracellular maturation of the ST3 zymogen suggested that the zymogen displays an encrypted domain that initiates processing. Amino-acid sequence alignments comparing ST3 with other MMPs have identified a non-homologous 10-amino-acid insert that is sandwiched between the pro-domain and the N terminus of the active proteinase¹ (Fig. 2a). To determine whether this decapeptide regulates processing, the effect of switching this domain from ST3 to the homologous region in a structurally distinct MMP, human fibroblast collagenase^{1,5} (HFC), was assessed. As expected, although wild-type ST3 was processed to its active form, HFC was secreted as glycosylated and non-glycosylated zymogens⁵ (Fig. 2a, c). After domain switching, however, the deletion mutant of ST3 (termed ST3⁻¹⁰) was not processed and secreted only as a single, inactive ~65K species (Fig. 2b, c). In contrast, the insertional mutant of HFC (termed HFC⁺¹⁰) was secreted as a fully processed and active

* To whom correspondence should be addressed.

FIG. 1 Activation and processing of ST3. **a**, Detection of ST3 activity. α_1 PI was incubated alone (lane 1), with supernatants from cells stably transfected with the control expression vector (lane 2), or the ST3 expression vector in the absence or presence of rTIMP-1 ($1 \mu\text{g ml}^{-1}$), rTIMP-2 ($1 \mu\text{g ml}^{-1}$) or BB-94 ($5 \mu\text{M}$) (lanes 3–6, respectively) and examined by SDS-PAGE/Coomassie staining. α_1 PI, α_1 -proteinase inhibitor; α_1 PI_x, cleaved α_1 PI. **b**, ST3 activation by transfected cell lines. COS, MCF-7 or HT1080 cells transiently transfected with α_1 PI and either the control (lanes 1, 3, 5) or the ST3 expression vector (lanes 2, 4, 6) were analysed by immunoprecipitation with ST3- or α_1 PI-specific polyclonal antisera (upper and lower panels, respectively). The 65K and 45K species are the zymogen and processed forms of ST3, respectively (upper panels). Cleaved α_1 PI is detected when coexpressed with ST3 in all three cell lines (lower panels; lanes 2, 4, 6) but not in control transfected cells (lanes 1, 3, 5). **c**, Pulse-chase analysis of ST3 processing *in situ*. COS-7 cells stably transfected with ST3 were pulse-labelled with [^{35}S]methionine for 5 min and chased with unlabelled methionine for 0 min (lanes 1 and 6), 20 min (lanes 2 and 7), 40 min (lanes 3 and 8), 60 min (lanes 4 and 9) and 80 min (lanes 5 and 10). Supernatants (lanes 1–5) and lysates (lanes 6–10) were immunoprecipitated with anti-ST3 polyclonal antisera. The arrow indicates the position of the non-glycosylated ST3 precursor.

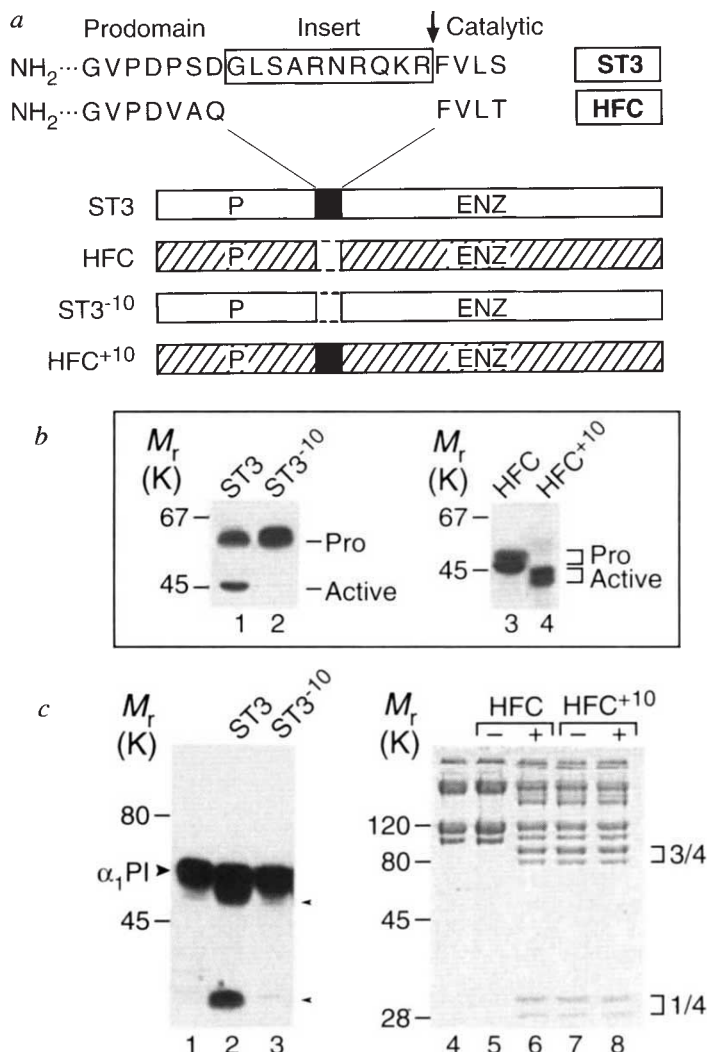
METHODS. COS-7 cells stably transfected with the control or ST3 expression vector have been described previously⁸. Serum-free conditioned media was collected from either cell population and incubated with α_1 PI ($10 \mu\text{g ml}^{-1}$) in the absence or presence of MMP inhibitors for 6 h at 37°C and analysed as described by SDS-PAGE (7.5%)⁸. COS-7, HT-1080 or MCF-7 cells were transiently transfected with the α_1 PI and ST3 expression vectors by LipofectAMINE treatment (BRL-GIBCO) and labelled with [^{35}S]methionine ($100 \mu\text{Ci ml}^{-1}$ for 3 h at 37°C). For pulse-chase analyses, stably transfected COS-7 cells were labelled with



[^{35}S]methionine ($500 \mu\text{Ci ml}^{-1}$) for 5 min and chased with 10 mM unlabelled methionine. Cells were lysed in RIPA buffer²³ in the presence of PMSF, E-64, BB-94 and pepstatin. ST3 and α_1 PI were immunoprecipitated with monospecific polyclonal antisera and processed as described⁸.

FIG. 2 A new 10 amino-acid insert in ST3 encodes an intracellular activation signal. **a**, Schematic representation of domain swaps between ST3 and HFC. The 10 amino-acid insert of ST3 (black box) is located between the pro- (labelled P) and catalytic (labelled ENZ) domains of ST3. ST3⁻¹⁰ is a deletion mutant lacking the insert between residue Asp 87 and Phe 98 whereas HFC⁺¹⁰ contains this motif inserted between residue Gln 99 and Phe 100 (ref. 5). **b**, Processing of ST3 and HFC mutants. ST3 (lane 1), but not ST3⁻¹⁰ (lane 2) was processed to its mature form. In contrast, HFC was secreted as glycosylated and non-glycosylated proforms⁵ (lane 3) whereas HFC⁺¹⁰ was processed completely to the mature enzyme (lane 4). Products were visualized by SDS-PAGE/fluorography after immunoprecipitation. Brackets indicate the glycosylated and non-glycosylated forms of HFC. **c**, Enzymic activities of ST3, ST3⁻¹⁰, HFC and HFC⁺¹⁰. Shown are α_1 PI expressed alone (lane 1) or α_1 PI coexpressed with ST3 (lane 2) or ST3⁻¹⁰ (lane 3). The cleaved α_1 PI products includes the ~50K fragment (upper arrowhead) and the ~4K fragment (lower arrowhead). Native type I collagen (lane 4) was incubated with HFC or HFC⁺¹⁰ (lanes 5 and 7, respectively) or aminophenylmercuric acetate-treated HFC or HFC⁺¹⁰ (lanes 6 and 8, respectively). Although the HFC zymogen expressed no collagenolytic activity until activated with the organomercurial, HFC⁺¹⁰ was secreted as the fully active enzyme. The characteristic 3/4- and 1/4-sized fragments of cleaved type I collagens are indicated.

METHODS. A sequential PCR strategy was used to generate both ST3⁻¹⁰ and HFC⁺¹⁰ (ref. 23). A 5' primer containing the ATG codon and a 3' primer containing the stop codon of ST3 were paired with two complementary internal primers encoding the deletion of the 10 amino-acid insert to generate two partial PCR fragments which were annealed and reamplified to generate ST3⁻¹⁰. HFC⁺¹⁰ was generated in a similar fashion with the following primers: a 5' and a 3' primer of HFC, and two complementary internal primers with a precise insertion of the 10 amino-acid insert of ST3 between Q 99 and F 100 of HFC. The ST⁻¹⁰ and HFC⁺¹⁰ PCR fragments were cloned and sequenced in pCloneAmp (BRL-GIBCO), and cloned into pREP9. DNA transfection and immunoprecipitation was as in Fig. 1a. For N-terminal sequence determination, [^3H]leucine-labelled HFC⁺¹⁰ was immunoprecipitated with specific polyclonal antisera, electrophoresed, immunoblotted and sequenced as described^{8,24}. Collagenolytic activity was analysed after a 12 h incubation with soluble type I collagen at 25°C as described²⁵.



type I collagenase with a new N terminus immediately downstream of the decapeptide at Phe 100 (Fig. 2b, c). Thus, despite the fact that HFC displays only limited homology to ST3¹, the 10 amino-acid insert carries all of the information necessary to direct the intracellular processing of the recipient MMP.

The ST3 decapeptide insert contains a tetrad of basic residues arranged in an Arg-X-Arg-X-Lys-Arg sequence (Fig. 2). Interestingly, recent studies indicate that protein precursors that display a triad of basic residues arranged in an Arg-X-Lys/Arg-Arg motif can be cleaved on the C-terminal side of the consensus sequence within the constitutive secretory pathway by mammalian homologues of the yeast processing protease, Kex2 (refs 6, 7). At least six members of the mammalian precursor processing endoproteases have been identified, but only two of these, furin and PACE4, are ubiquitously expressed and display processing activities for constitutively secreted, Arg-X-Lys/Arg-Arg-containing precursors^{6,7}. Sequence rules established for precursor cleavage by furin/PACE4-like convertases indicate critical roles for the basic residues at positions -1, -2 and -4 relative to the scissile bond (that is, P⁻¹, P⁻² and P⁻⁴, respectively)^{10,11}. Thus, the amino-acid sequence requirements for ST3 processing in COS-7 cells¹² were compared in transient transfection assays where the Arg residues at P⁻⁴, P⁻² and P⁻¹ were substituted. In contrast to wild type ST3, Arg→Lys or

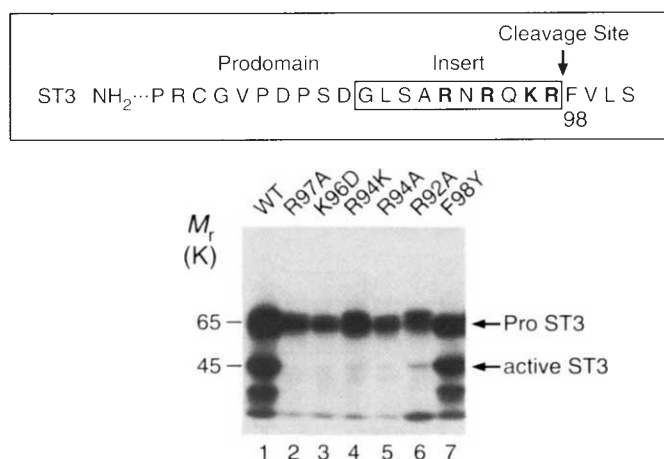


FIG. 3 Amino-acid sequence requirements for ST3 processing. The basic residue motif in the 10 amino-acid insert (boxed) of ST3 is shown in bold letters. Mutations were introduced in each of these residues as well as the Phe 98 by a sequential PCR-based strategy. ST3 expression constructs harbouring these mutations were transiently transfected into COS-7 cells and analysed after immunoprecipitation and SDS-PAGE/fluorography as described in Fig. 1.

METHODS. Mutations were introduced into the desired positions in ST3 essentially as described for the generation of ST⁻¹⁰ in Fig. 2. Mutagenic primers used are as follows: R97A, CGCAACCGACAGAAGCGTTCGTGCTTTCTGGC; K96D, GCCCGCAACCGACAGGATAGGTTCTGCTTTCT; R94K, CTGAGTGCCCGCAACAGCAGAAGAGGTTCTGTG; R94A, CTGAGTGCCCGCAACCGACAGAAGAGGTTCTGTG; R92A, GATGGCTGAGTGCCCGCAACCGACAGAAGAGGTTCTGTG; F98Y, AACCGACAGAAGAGGTATGTGCTTTCTGGCGGG. Bold nucleotides indicate the altered codons. These mutagenic primers were paired with the ST3 3' primer to generate C-terminal ST3 fragments carrying the desired mutations by PCR from an ST3 cDNA template⁸. A PCR fragment coding for the pro-domain of ST3 (M1-K 96) was generated by amplifying the ST3 cDNA template with the 5' ST3 primer described previously and the primer, CGCTTCTGTGGTTCGCGGCACTACGCCATC, whose complementary strand encodes D 87-K 96. The M1-K 96 fragment was annealed to each of the mutated C-terminal ST3 fragment generated above and PCR amplified with the 5' and 3' ST3 primers to generate full-length ST3 mutants as described in Fig. 2. The resulting PCR fragments were cloned into pCloneAmp, screened for mutations by DNA sequencing, and cloned into pREP9 expression vector.

Arg→Ala mutants at P⁻⁴ (R94K and R94A, respectively), Lys→Asp at P⁻² (K96D) or Arg→Ala at P⁻¹ (R97A), were all secreted as unprocessed zymogens (Fig. 3). Although an additional basic residue at P⁻⁶ has not been identified in other Arg-X-Lys/Arg-Arg-containing precursors, mutants containing this extended recognition sequence have been reported to undergo more efficient processing^{10,11,13}. Indeed, an Arg→Ala substitution at P⁻⁶ was processed, but only at a rate ~10% of that

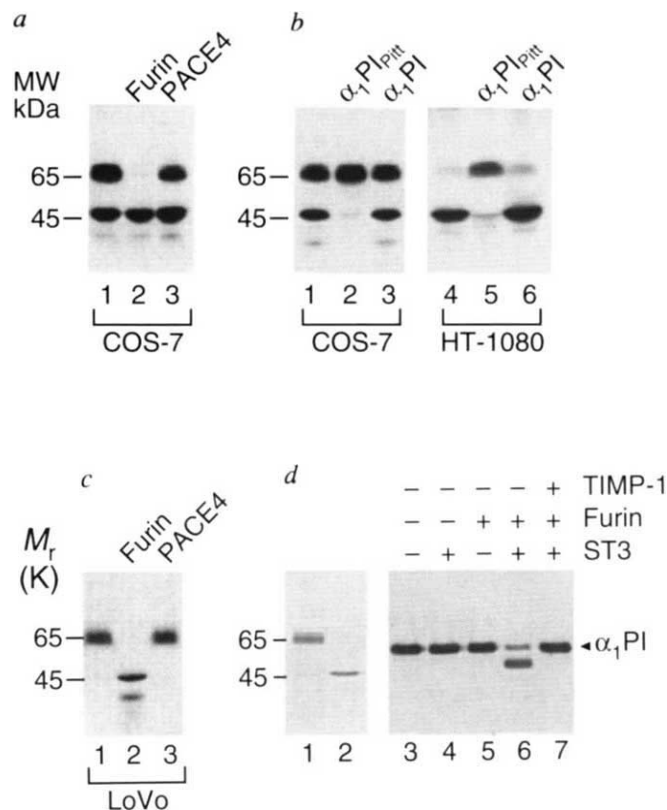


FIG. 4 Processing of ST3 by furin and PACE4. **a**, Efficiency of furin and PACE4 as ST3-processing convertases. ST3 expression vector (1 μ g) was cotransfected with 100 ng control vector (lane 1) or expression vectors for human furin (lane 2) or PACE4 (lane 3) into COS-7 cells. ST3 products were analysed after immunoprecipitation as described in Fig. 1. **b**, Inhibition of ST3 activation by α_1 PI_{in situ}. COS-7 cells (lanes 1–3) or HT-1080 cells (lanes 4–6) were transfected with the following expression vectors; ST3 alone (1 μ g, lanes 1 and 4), ST3 and α_1 PI_{in situ} (1 μ g and 100 ng, respectively; lanes 2 and 5), or ST3 and α_1 PI (1 μ g and 100 ng, respectively; lanes 3 and 6). The ST3 products were analysed following immunoprecipitation as described in Fig. 1. **c**, ST3 processing in the furin-deficient cell line, LoVo. LoVo cells were transfected with expression vector for ST3 (lane 1), or with expression vectors for ST3 and either furin (1 μ g; lane 2) or PACE4 (1 μ g; lane 3). Radiolabelled ST3 products were analysed after immunoprecipitation as described in Fig. 1. **d**, Processing of proST3 by soluble furin under cell-free conditions. ProST3 (50 ng) was then incubated alone (lane 1) or with purified soluble furin (2 units; lane 2) and the mixtures analysed by western blotting with ST3 polyclonal antisera. To assess ST3 proteolytic activity, α_1 PI (200 ng) was incubated alone (lane 3) or with proST3 (20 ng; lane 4), soluble furin (2 units; lane 5), furin-activated ST3 (lane 6) or furin-activated ST3 and TIMP-1 (50 ng; lane 7). Products were analysed by western blotting with α_1 PI polyclonal antisera.

METHODS. ProST3 was purified from transiently transfected LoVo cells as described⁸. Soluble furin expressed in transiently transfected COS-7 cells was purified and assayed according to ref. 26. Furin-ST3 mixtures were incubated for 2 h at 37 °C in 10 mM Tris-HCl, 250 mM NaCl, 1 mM CaCl₂ and 0.02% Brij 35 (pH 7.0). N-terminal sequence analysis of furin-processed ST3 was determined as described⁸ in the absence or presence of BB-94. α_1 PI hydrolysis was determined under conditions in Fig. 1 and the antipeptidase immunoblotted with polyclonal antisera (Calbiochem) as described⁸.

observed with the wild-type proteinase (Fig. 3). Finally, given that substitutions in the P⁺ position are more readily tolerated^{10,11}, F98Y was processed comparably to wild-type ST3 (Fig. 3). Under the assumption that insert substitutions did not significantly perturb the conformation of the Arg 97-Phe 98 cleavage site^{12,14}, these results suggest that ST3 undergoes furin- or PACE4-mediated processing.

To assess the relative efficiency of furin versus PACE4 in ST3 processing, COS-7 cells were cotransfected with ST3 cDNA and either furin or PACE4 cDNAs (Fig. 4a). Although COS cells transfected with either pro-protein convertase express the respective proteases comparably¹², only furin increased ST3 processing (Fig. 4a). Furthermore, when ST3-transfected COS-7 or HT-1080 cells were cotransfected with the Pittsburgh mutant of α_1 PI (α_1 PI_{Pitt}), a reactive site variant that inhibits furin (but not PACE4) activity *in situ*^{12,14}, ST3 processing was completely blocked (Fig. 4b). Consistent with these findings, LoVo cells, a carcinoma cell line that does not produce functional furin¹⁵, were unable to process the ST3 zymogen to its active form (Fig. 4c). However, when LoVo cells were cotransfected with ST3 and furin (but not PACE4) cDNAs, processing was reestablished (Fig. 4c). Finally, to determine whether furin directly mediates ST3 processing, a soluble form of the convertase was generated by deleting the transmembrane domain^{12,14} and the purified mutant incubated with the ST3 zymogen under cell-free conditions. At neutral pH, the soluble furin mutant efficiently cleaved the ST3 zymogen at the Arg 97-Phe 98 junction (as determined by N-terminal sequencing) to generate the active 45K form of the proteinase (Fig. 4d).

We have demonstrated that ST3 is processed directly to its mature active form by the pro-protein convertase, furin. A mammalian homologue of the yeast Kex2 pheromone convertase, furin is a transmembrane serine proteinase concentrated in the *trans*-Golgi network^{6,7,16,17}. Furin efficiently processes serum proteins, growth factors and membrane receptors containing an Arg-X-Lys-Arg sequence^{6,7}, but an additional Arg residue at P⁻⁶ was necessary for the efficient processing of the ST3 zymogen. Apparently, the inclusion of this enhancing signal for human ST3 processing is evolutionarily conserved because mouse² and *Xenopus*¹⁸ ST3 homologues each contain the identical motif. Importantly, a new transmembrane MMP that controls gelatinase A activation also contains a decapeptide insert with an Arg-X-Lys-Arg motif upstream of its catalytic domain^{19,20}. Thus, we propose that pro-protein convertases may serve as important regulators of multiple metalloproteinases involved in extracellular matrix turnover. The unexpected identification of furin as an activator of MMPs that are strongly implicated in tumour progression suggests that recently developed pro-protein convertase inhibitors^{21,22,27} might be developed as anti-cancer therapeutics. □

19. Sato, H. *et al.* *Nature* **370**, 61–65 (1994).
20. Cao, J., Sato, H., Takino, T. & Seiki, M. *J. biol. Chem.* **270**, 801–805 (1995).
21. Hallenberger, S. *et al.* *Nature* **360**, 358–361 (1992).
22. Decroly, E. *et al.* *J. biol. Chem.* **269**, 12240–12247 (1994).
23. Ausubel, F. M. *et al.* (eds) *Curr. Prot. molec. Biol.* (Wiley, New York, 1994).
24. Birken, S., Fetherston, J., Canfield, R. & Boime, R. *J. biol. Chem.* **256**, 1816–1823 (1981).
25. Desrochers, P. E., Mookhtiar, K., Van Wart, H., Hasty, K. A. & Weiss, S. J. *J. biol. Chem.* **267**, 5005–5012 (1992).
26. Bravo, D. A., Gleason, J. B., Sanchez, R. I., Roth, R. A. & Fuller, R. S. *J. biol. Chem.* **269**, 25830–25837 (1994).
27. Anderson, E. A., Thomas, L., Hayflick, J. S. & Thomas, G. *J. biol. Chem.* **268**, 24887–24891 (1993).

ACKNOWLEDGEMENTS. We thank A. Rehemtulla and R. Kaufman (HHMI, University of Michigan) for pro-peptide convertase, α_1 PI, α_1 PI_{Pitt}, and soluble furin expression vectors as well as helpful discussions. R. Fuller (University of Michigan) for guidance in purifying soluble furin. A. Galloway (British Biotechnology) for BB-94, C. Thompson (Synergen) for rTIMP-1, K. Langley (Amgen) for rTIMP-2 and H. Welgus for HFC polyclonal antisera. D.P. was supported in part by an Oncology Research Training Grant from the National Heart, Lung and Blood Institute. This study was supported by grants from the University of Michigan Cancer Research Committee, the Michigan Breast Care Center, the Charlotte Geyer Foundation and the US Army Medical Research Command.

Direct stimulation of Jak/STAT pathway by the angiotensin II AT₁ receptor

Mario B. Marrero*, Bernhard Schieffer*, William G. Paxton*, Lauri Heerdt†, Bradford C. Berk‡, Patrick Delafontaine§ & Kenneth E. Bernstein*||

Departments of * Pathology and Laboratory Medicine, and § Medicine, Division of Cardiology, Emory University, Atlanta, Georgia 30322, USA

† Santa Cruz Biotechnology Inc., Santa Cruz, California 95060, USA

‡ Department of Medicine, Division of Cardiology, University of Washington, Seattle, Washington 98195, USA

THE peptide angiotensin II is the effector molecule of the renin-angiotensin system. All the haemodynamic effects of angiotensin II, including vasoconstriction and adrenal aldosterone release, are mediated through a single class of cell-surface receptors known as AT₁ (refs 1, 2). These receptors contain the structural features of the G-protein-coupled receptor superfamily³. We show here that angiotensin II induces the rapid phosphorylation of tyrosine in the intracellular kinases Jak2 and Tyk2 in rat aortic smooth-muscle cells and that this phosphorylation is associated with increased activity of Jak2. The Jak family substrates STAT1 and STAT2 (for signal transducers and activators of transcription) are rapidly tyrosine-phosphorylated in response to angiotensin II. We also find that Jak2 co-precipitates with the AT₁ receptor, indicating that G-protein-coupled receptors may be able to signal through the intracellular phosphorylation pathways used by cytokine receptors^{4,5}.

To investigate whether angiotensin II (Ang II) stimulates Jak2 phosphorylation, rat aortic smooth-muscle (RASM) cells were exposed to Ang II, and Jak2 tyrosine-phosphorylation was measured by two methods. In the first, cell lysates were immunoprecipitated with a monoclonal anti-phosphotyrosine antibody, precipitated proteins were separated by gel electrophoresis, transferred to nitrocellulose and then immunoblotted with a polyclonal anti-Jak2 antibody (Fig. 1a). In the second protocol, the order of antibody addition was reversed: the cell lysate was immunoprecipitated with anti-Jak2 and then immunoblotted with anti-phosphotyrosine (Fig. 1b). Similar experiments were done in parallel in which RASM cells were exposed to murine interferon (IFN)- γ , a cytokine that activates phosphorylation of Jak2 (Fig. 1c)^{4,5}. Within 5 min of exposure to Ang II, tyrosine

Received 30 January; accepted 22 March 1995.

1. Basset, P. *et al.* *Nature* **348**, 699–704 (1990).
2. Lefebvre, O. *et al.* *J. Cell Biol.* **119**, 997–1002 (1992).
3. Wolf, C. *et al.* *Proc. natn. Acad. Sci., U.S.A.* **90**, 1843–1847 (1993).
4. Matrisian, L. M. *BioEssays* **14**, 455–463 (1992).
5. Birkedal-Hansen, H. *et al.* *Crit. Rev. Oral Biol. Med.* **4**, 197–250 (1993).
6. Steiner, D. F., Smeekens, S. P., Ohagi, S. & Chan, S. J. *J. biol. chem.* **267**, 23435–23438 (1992).
7. Seidah, N. G., Day, R. & Chretien, M. *Biochem. Soc. Trans.* **21**, 685–691 (1993).
8. Pei, D. Q., Majumdar, G. & Weiss, S. J. *J. biol. Chem.* **269**, 25849–25855 (1994).
9. Davies, B., Brown, P. Q., East, N., Crimmin, M. J. & Balkwill, E. R. *Cancer Res.* **53**, 2087–2091 (1993).
10. Watanabe, T. *et al.* *J. biol. Chem.* **267**, 8270–8274 (1991).
11. Watanabe, T., Murakami, K. & Nakayama, K. *FEBS Lett.* **320**, 215–218 (1993).
12. Rehemtulla, A. & Kaufman, R. *J. Blood* **79**, 2349–2355 (1992).
13. Liu, Y.-C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **90**, 8957–8961 (1993).
14. Wasley, C. L., Rehemtulla, A. & Kaufman, R. J. *J. biol. Chem.* **268**, 8458–8465 (1993).
15. Tsuneoka, M. *et al.* *J. biol. Chem.* **268**, 26461–26465 (1993).
16. Molloy, S. S., Thomas, K. J., Van Slyke, K., Stenberg, P. E. & Thomas, G. *EMBO J.* **13**, 18–33 (1994).
17. Vidricaire, J., Denault, J.-B. & Leduc, R. *Biochem. biophys. Res. Commun.* **195**, 1011–1018 (1993).
18. Patterson, D., Hayes, W. P. & Shi, Y. B. *Dev. Biol.* **167**, 252–262 (1995).

To whom correspondence should be addressed.

in Jak2 was phosphorylated, with IFN- γ stimulating Jak2 tyrosine-phosphorylation as expected. The temporal response of Ang II- and IFN- γ -treated cells was different. In cells exposed to Ang II, Jak2 tyrosine-phosphorylation returned to control levels by 60 min; phosphorylation in response to IFN- γ remained high at the 60-min point. In separate experiments, RASM cells were exposed to epidermal growth factor (EGF) or to IFN- α ; no increase in Jak2 phosphorylation on tyrosine was detected, consistent with previous reports for these two cytokines^{5,6}.

We used a similar protocol to investigate the Ang II-mediated tyrosine-phosphorylation of Jak1 and Tyk2, two other members of the Jak family of intracellular kinases. RASM cells were exposed to Ang II and cell lysates were immunoprecipitated with anti-phosphotyrosine antiserum; these proteins were then immunoblotted with either a rabbit polyclonal anti-Jak1 or anti-Tyk2 antibody (Fig. 1d,f). In parallel experiments, RASM cells were exposed to murine IFN- α , which stimulates tyrosine-phosphorylation of Jak1 and Tyk2 (Fig. 1e,g)^{5,7}. We found that Tyk2 was tyrosine-phosphorylated in response to Ang II but that

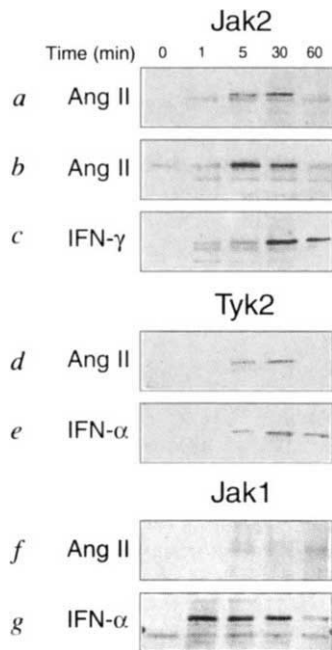


FIG. 1 Tyrosine-phosphorylation of Jak2, Tyk2 and Jak1 in response to Ang II. Cells were stimulated for 0, 1, 5, 30 or 60 min with Ang II (a, b, d, f), IFN- γ (c) or IFN- α (e, g). All cell lysates were immunoprecipitated with anti-phosphotyrosine antibodies except b, which was precipitated with anti-Jak2. Western blots of the precipitated proteins were then probed with anti-Jak2 (a, c), anti-phosphotyrosine (b), anti-Tyk2 (d, e), or anti-Jak1 (f, g). Ang II stimulated the tyrosine-phosphorylation of Jak2 and Tyk2 but not Jak1. All experiments were done at least three times. METHODS. RASM cells were serum-starved for 24 to 36 h, and treated for the indicated times with either Ang II (10^{-7} M), IFN- γ (10 ng ml $^{-1}$), or IFN- α ($1,000$ U ml $^{-1}$). Cells were lysed in 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2.5 mM EDTA, 1.0% Triton X-100, 10% glycerol, 0.1% SDS, 50 mM NaF, 10 mM Na $_3$ P $_2$ O $_7$, 1.0% deoxycholic acid, 1 mM Na $_3$ VO $_4$, 1 mM PMSF and 10 μ g ml $^{-1}$ aprotinin. After clarification by centrifugation, lysates were immunoprecipitated, collected on Protein A/G agarose, washed with 50 mM Tris, pH 8.0, 150 mM NaCl, 0.1% Triton X-100 and eluted with 2 $\times$ Laemmli's sample buffer²⁶. Proteins were separated by SDS-PAGE and transferred to nitrocellulose. Blots were probed as described above and developed with enhanced chemiluminescence (Amersham) as described by the manufacturer. Monoclonal anti-phosphotyrosine antibodies were from Upstate Biotechnology Inc. and Transduction Laboratories; polyclonal anti-Jak1, anti-Jak2 and anti-Tyk2 antibodies were from Santa Cruz Biotechnology Inc.

Jak1 was not. Jak1 tyrosine-phosphorylation was stimulated by IFN- α as well as by IFN- γ (data not shown).

To determine whether Jak2 kinase activity was stimulated by Ang II, we measured enzyme autophosphorylation *in vitro* (Fig. 2)⁸. RASM cells were submaximally stimulated with Ang II to obtain limited tyrosine-phosphorylation of Jak2, which was then immunoprecipitated and reacted with ATP to permit autophosphorylation. Phosphorylation on tyrosine was quantified by western blot analysis using an anti-phosphotyrosine antibody, which showed that Ang II induces Jak2 autophosphorylation.

To test whether Jak2 was associated with the AT $_1$ receptor, we generated a rabbit polyclonal antiserum against a glutathione-S-transferase (GST)-fusion protein containing amino acids 306–359 of the rat AT $_{1A}$ receptor³, which constitute the carboxy-terminal 54 amino acids of this receptor. Western blot analysis with this antiserum revealed a major band corresponding to M_r 67K and a minor band of 72K. These bands remain when the antiserum is preabsorbed with GST, but disappear when the antiserum is preabsorbed with the GST-fusion protein (data not shown). Cell lysates from RASM cells treated with Ang II were immunoprecipitated with the anti-receptor antiserum and the precipitated proteins probed with anti-Jak2 antibody (Fig. 3A). The experiment shows that Jak2 co-precipitates with the AT $_1$ receptor. Although small amounts of Jak2 were occasionally bound to the AT $_1$ receptor in unstimulated cells, addition of Ang II increased the AT $_1$ -receptor–Jak2 association within 1 min. In parallel experiments, RASM cell lysates were immunoprecipitated with either a rabbit polyclonal anti-GST antiserum or rabbit preimmune serum: these control sera did not co-precipitate with Jak2. We also tested Jak1 or Tyk2 for co-precipitation with the AT $_1$ receptor using commercially available antibodies, but found that neither of these kinases associated with the AT $_1$ receptor in response to Ang II.

We next investigated whether Ang II-mediated phosphorylation of Jak2 could be the work of a secreted cytokine. RASM cells were exposed to Ang II, and cell supernatants containing Ang II peptide and any potential cytokines were collected. After

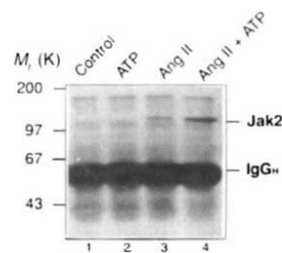


FIG. 2 Jak2 tyrosine kinase activity measured by autophosphorylation. RASM cells were incubated with 10^{-9} M Ang II for 30 min at 25 °C to stimulate limited receptor activation and partial Jak2 autophosphorylation. Jak2 was then immunoprecipitated and allowed to autophosphorylate *in vitro* in the presence of exogenous ATP (lane 4). Tyrosine phosphorylation was measured by immunoblotting with anti-phosphotyrosine antibody. Control conditions are shown in lanes 1–3: 1, cells were not exposed to Ang II nor was exogenous ATP added to the *in vitro* autophosphorylation reaction; 2, cells were not exposed to Ang II but ATP was added to the *in vitro* autophosphorylation reaction; 3, cells were exposed to Ang II but no exogenous ATP was added during the *in vitro* autophosphorylation step. The small signal seen in this lane is probably Jak2 autophosphorylation with endogenous ATP.

METHODS. Jak2 immunoprecipitates were immobilized on Protein A/G agarose beads and washed three times in kinase buffer (50 mM HEPES, pH 7.3, 100 mM NaCl, 0.2 mM Na $_3$ VO $_4$, 0.1% Triton X-100, 1 mM PMSF and 10 μ g ml $^{-1}$ aprotinin). The beads were resuspended in a 100 μ l kinase buffer with 3 mM MnCl $_2$ and 15 μ M ATP. After incubation, samples were washed in kinase buffer, separated by SDS-PAGE, transferred to nitrocellulose and probed with anti-phosphotyrosine antibody. The blot was developed with an alkaline phosphatase-conjugated goat anti-mouse antibody (BioRad) as described²⁰.

addition of losartan (Dup753), a specific non-peptide inhibitor of the AT₁ receptor, we tested the supernatants for their ability to stimulate Jak2 phosphorylation (Fig. 3B) and found that whereas Ang II stimulated tyrosine phosphorylation of Jak2, the supernatants of Ang II-treated cells did not. The results of this experiment and of the AT₁-Jak2 co-precipitation indicate that the Ang II AT₁ receptor directly mediates tyrosine-phosphorylation of Jak2.

In cytokine-responsive cells, stimulation of Jak activity leads to phosphorylation of STAT proteins⁵⁻⁷. To see whether there is a similar effect in cells exposed to Ang II, we immunoprecipitated cell lysates with a monoclonal anti-phosphotyrosine antibody and immunoblotted with antibodies against p91/84 (STAT1 α/β), p113 (STAT2) or p92 (STAT3) (Fig. 4A, a, c and e). In parallel, cells were exposed to IFN- γ , IFN- α or EGF (Fig. 4A, b, d, f). Ang II was found to induce increased tyrosine-phosphorylation of all the STAT proteins: tyrosine-phosphorylation of STAT1 α/β was increased within 5 min of Ang II addition and was maximal at 5 or 15 min, depending on the

experiment. STAT2 showed a similar pattern of tyrosine-phosphorylation, but none was detected in STAT3 until 60 min after addition of Ang II. As expected, IFN- γ stimulated tyrosine-phosphorylation of STAT1 α/β and IFN- α of STAT2. Neither IFN- γ nor EGF induced tyrosine-phosphorylation of STAT2, but EGF did increase STAT3 tyrosine-phosphorylation at 15 min; there was no increase in STAT3 tyrosine-phosphorylation in response to IFN- α or IFN- γ .

When cells are exposed to an interferon, STAT proteins are phosphorylated and assembled into a multiprotein complex which translocates to the nucleus⁶. We tested whether Ang II could induce nuclear translocation of STAT1 by exposing RASM cells to Ang II and probing nuclear proteins by western

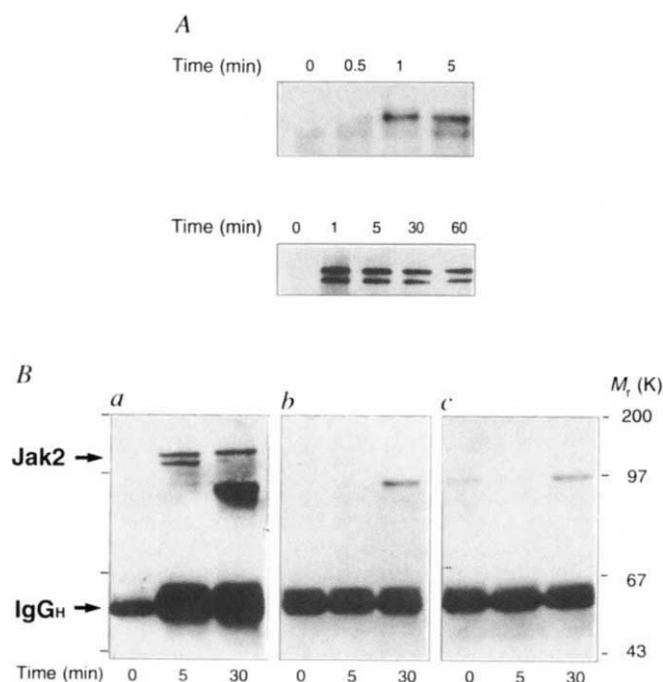


FIG. 3 A, Time course of Ang II-induced association of Jak2 with the AT₁ receptor. RASM cells were stimulated with 10^{-7} M Ang II for the indicated times (min). Cell lysates were immunoprecipitated with anti-AT₁ receptor antibody (Santa Cruz Biotechnology Inc.). Western blots of the immunoprecipitates were then probed with anti-Jak2 antibody. Results shown in the two panels are from two separate experiments; in total the experiment was done three times. B, Ang II stimulates Jak2 phosphorylation via the AT₁ receptor. In a, RASM cells were stimulated for 0, 5 or 30 min with Ang II (10^{-7} M). Cells were then lysed, and Jak2 phosphorylation assessed as for Fig. 1. In b, the medium from cells treated for 5 min with Ang II in a was collected and losartan, a specific and competitive AT₁-receptor inhibitor, was added at 10^{-5} M (ref. 27). The medium was then added to RASM cells for the indicated times, and Jak2 phosphorylation assessed as for Fig. 1. In c, the medium from cells treated for 30 min with Ang II (a) was collected. Losartan was added and tyrosine-phosphorylation of Jak2 assessed as in b, but no tyrosine-phosphorylation was observed. Losartan also blocked the Ang II-mediated phosphorylation of Jak2 seen in a (data not shown). These results show that the action of Ang II is direct and not the result of a secreted cytokine. METHODS. In A, serum-starved RASM cells were treated with 10^{-7} M Ang II as indicated. Cells were lysed in a buffer composed of 1% Triton X-100, 20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2.5 mM EDTA, 1 mM Na₃VO₄, 1 mM PMSF and 10 μ g ml⁻¹ aprotinin. Immunoprecipitation and western blotting are described in Fig. 1 legend.

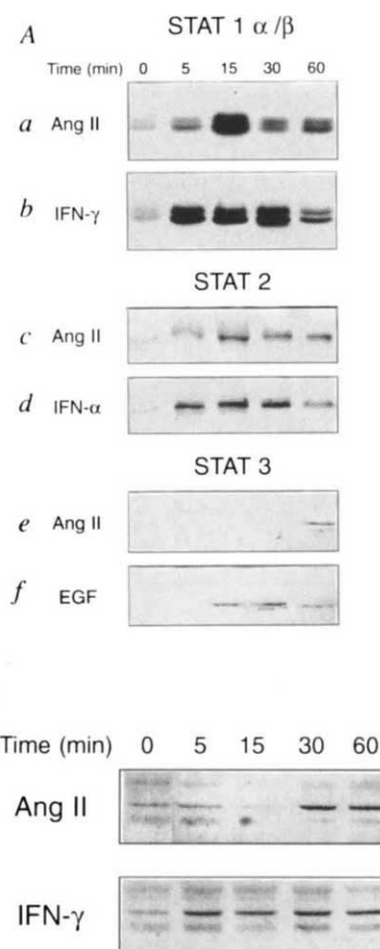


FIG. 4 A, Tyrosine phosphorylation of STAT1 α/β , STAT2 and STAT3 in response to angiotensin II. Cells were stimulated for the indicated times with Ang II (a, c, e), IFN- γ (b), IFN- α (d) or EGF (f). Western blots of precipitated proteins were then probed with anti-STAT1 α/β (a, b), anti-STAT2 (c, d) or anti-STAT3 (e, f). In a and b, the upper band is p91 and the lower band is p84. B, Nuclear translocation of STAT1 in response to Ang II. RASM cells were stimulated for the indicated times with either Ang II (top panel) or IFN- γ (bottom panel). Nuclear proteins were then prepared and western blots probed with anti-STAT1 α/β antibody. METHODS. For A, serum-starved RASM cells were treated with Ang II (10^{-7} M), IFN- γ (10 ng ml⁻¹), IFN- α ($1,000$ U ml⁻¹) or EGF (100 ng ml⁻¹). Cells were lysed as for Fig. 1, immunoprecipitated with anti-phosphotyrosine antibody and western blotted. Anti-phosphotyrosine antibodies were from Transduction Laboratories; polyclonal anti-STAT1 α/β , STAT2 and STAT3 antibodies were from Santa Cruz Biotechnology Inc. For B, serum-starved RASM cells were treated for the indicated times with Ang II (10^{-7} M) or IFN- γ (10 ng ml⁻¹). Nuclear extracts were prepared as described²⁸. Total nuclear protein (30 μ g) was separated by SDS-PAGE, transferred to nitrocellulose, immunoblotted with anti-STAT1 α/β and detected by alkaline phosphatase as described in Fig. 2 legend.

blotting with anti-STAT1 antibody; in a parallel experiment, cells were exposed to IFN- γ . We found that Ang II induced nuclear translocation of STAT1 30 min after ligand addition, whereas it took IFN- γ less than 5 min (Fig. 4B). A similar result has been reported in neonatal rat cardiac fibroblasts⁹.

Many cell-surface cytokine receptors are known to trigger intracellular phosphorylation cascades^{10,11}, either as a result of their own intrinsic kinase activity or through their association with intracellular kinases¹²⁻¹⁷. Intracellular phosphorylation initiated by receptors of the seven-helix transmembrane type such as AT₁ is achieved through interaction with GTP-binding proteins, but there is evidence to suggest that kinase activation also plays an important role in the signalling pathway^{18,19}. In RASM cells, for instance, phospholipase C- γ 1 is phosphorylated on tyrosine within 30 s of exposure to Ang II (ref. 20), and in rat neonatal cardiac fibroblasts Ang II induces rapid tyrosine-phosphorylation of several proteins, including p46/56^{SHC} (ref. 21), indicating that tyrosine kinase is activated in response to Ang II. We have shown that members of the Jak family of intracellular tyrosine kinases and their STAT1 and 2 substrates are rapidly phosphorylated on tyrosine and activated in response to Ang II. Jak2 seems to associate physically with the AT₁ receptor, indicating that Jak2 activation in response to ligand occupancy of the AT₁ receptor is analogous to cytokine receptor activation. Our results show that there are common features in the signalling cascades in RASM cells for such diverse agents as Ang II, EGF and the interferons. Angiotensin II has been shown to induce proto-oncogenes such as *c-fos* and to promote the growth of smooth muscle cells^{22,23}. In other cells the Jak-STAT pathway has been proposed as the signalling pathway used by cell-surface-binding cytokines that is responsible for transcriptional activation of early growth-response genes^{24,25}. The Jak-STAT pathway may thus play a similar role in the control of Ang II-mediated

cell growth. Our results identify a molecular mechanism that may underlie Ang II-induced growth responses in vascular smooth muscle cells. □

Received 16 February; accepted 4 April 1995.

1. Peach, M. J. *Physiol. Rev.* **57**, 313-337 (1977).
2. Peach, M. J. *Biochem. Pharmacol.* **30**, 2745-2751 (1981).
3. Murphy, T. J., Alexander, R. W., Griendling, K. K., Runge, M. S. & Bernstein, K. E. *Nature* **351**, 233-236 (1991).
4. Watling, D. et al. *Nature* **366**, 166-170 (1993).
5. Silvennoinen, O., Ihle, J. N., Schlessinger, J. & Levy, D. E. *Nature* **366**, 583-585 (1993).
6. Shuai, K. et al. *Nature* **366**, 580-583 (1993).
7. Velazquez, L., Fellous, M., Stark, G. R. & Pellegrini, S. *Cell* **70**, 313-322 (1992).
8. Rui, H., Kirken, R. & Farrar, W. L. *J. Biol. Chem.* **269**, 5364-5368 (1994).
9. Bhat, G. J., Thekkumkara, T. J., Thomas, W. G., Conrad, K. M. & Baker, K. M. *J. Biol. Chem.* **269**, 31443-31449 (1994).
10. Cohen, P. *Nature* **296**, 613-617 (1982).
11. Pazin, M. J. & Williams, L. T. *Trends Biochem. Sci.* **17**, 374-378 (1992).
12. Fantl, W. & Williams, L. T. *Rev. Biochem. Chem.* **62**, 453-481 (1993).
13. Schlessinger, J. & Ullrich, A. *Neuron* **9**, 383-391 (1992).
14. Cadena, D. L. & Gill, G. N. *FASEB J.* **6**, 2332-2337 (1992).
15. Silvennoinen, O. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8429-8433 (1993).
16. Miura, O. et al. *Blood* **84**, 1501-1507 (1994).
17. Keegan, A. D. et al. *Cell* **76**, 811-820 (1994).
18. Savarese, T. M. & Fraser, C. M. *Biochem. J.* **283**, 1-19 (1992).
19. Dohman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. *Rev. Biochem. Chem.* **60**, 651-688 (1992).
20. Marrero, M. B., Paxton, W. G., Duff, J. L., Berk, B. C. & Bernstein, K. E. *J. Biol. Chem.* **269**, 10935-10939 (1994).
21. Schorb, W., Peeler, T. C., Madigan, N. M., Conrad, K. M. & Baker, K. M. *J. Biol. Chem.* **269**, 19626-19632 (1994).
22. Geisterfer, A. A. T., Peach, M. J. & Owens, G. K. *Circ. Res.* **62**, 749-756 (1988).
23. Naftilan, A. J., Pratt, R. E. & Dzau, V. J. *J. Clin. Invest.* **83**, 1419-1424 (1989).
24. Darnell, J. E., Jr, Kerr, I. A. & Stark, G. R. *Science* **264**, 1415-1421 (1994).
25. Ihle, J. N. et al. *Trends Biochem. Sci.* **19**, 222-227 (1994).
26. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
27. Timmermans, P. B., Wong, P. C., Chiu, A. T. & Herblin, W. F. *Trends Pharmacol. Sci.* **12**, 55-62 (1991).
28. Dignam, J. D., Lebovitz, R. M. & Roeder, R. *Nucleic Acids Res.* **11**, 1475-1489 (1983).

ACKNOWLEDGEMENTS. We thank K. Frenzel, Q. Chai, M. Bencherif and H. Drexler. This work was supported by grants from the NIH, by Grants-in-Aid from the American Heart Association, and by the Georgia affiliate of the American Heart Association. M.B.M. is an American Heart Association Minority Development Award Scientist. B.S. was supported by the Deutsche Forschungsgemeinschaft. K.E.B., B.C.B. and P.D. are Established Investigators of the American Heart Association.

Specificity in chaperonin-mediated protein folding

Guoling Tian, Irina E. Vainberg, William D. Tap, Sally A. Lewis & Nicholas J. Cowan*

Department of Biochemistry, New York University Medical Center, New York, New York 10016, USA

CHAPERONINS are ubiquitous multisubunit toroidal complexes that aid protein folding in an ATP-dependent manner¹⁻⁶. Current models of folding by the bacterial chaperonin GroEL depict its role as unfolding and releasing molecules that have misfolded, so that they can return to a potentially productive folding pathway in solution^{7,8}. Accordingly, a given target polypeptide might require several cycles of binding and ATP-driven release from different chaperonin complexes before reaching the native state. Surprisingly, cycling of a target protein does not guarantee its folding, and we report here that unfolded β -actin or α -tubulin both form tight complexes when presented to either GroEL or its mitochondrial homologue, and both undergo cycles of release and rebinding upon incubation with ATP, but no native protein is produced. We conclude that different chaperonins produce distinctive spectra of folding intermediates.

To investigate whether cytosolic chaperonin (c-cpn, the homologue of GroEL in the eukaryotic cytosol^{5,9-15}) cycles its target proteins, we did a dilution experiment, reasoning that dilution

should reduce the efficiency with which released non-native forms can be recaptured by chaperonin. We found that dilution of a β -actin/c-cpn binary complex followed by incubation in the absence of ATP did not affect the level of binary complex (Fig. 1a). However, dilution in the presence of ATP resulted in an 85% reduction in the efficiency of native β -actin production compared with an undiluted control (Fig. 1b, c), with a progressive overall loss of radioactivity, presumably occurring because released non-native forms either aggregate or adhere to the walls of the reaction vessel. We also found that inclusion of excess mitochondrial chaperonin (mt-cpn) as a trap for the sequestration of non-native forms¹⁶ resulted in a decreased yield of native product similar to that found in our dilution experiments (Fig. 1d). These data are consistent with the accumulation of native protein as a result of several cycles of binding and release from c-cpn. The half-time for release of β -actin from c-cpn is about 5 min, whereas the half-time for generation of native product is 15-30 min. In both these respects, c-cpn is 5-10 times slower than GroEL/GroES^{7,8}; but the rate of ATP hydrolysis by c-cpn¹⁶ is about 3 times lower than the corresponding rate for GroEL/GroES⁷. Thus, c-cpn appears to function in a qualitatively similar manner to GroEL, but it does so more slowly.

To compare the ability of different chaperonins to bind target proteins, we presented an unfractionated mixture of labelled, unfolded *Escherichia coli* polypeptides to GroEL, mt-cpn or c-cpn (Fig. 2a). The pattern of molecules recognized by each chaperonin is quite similar, although the overall amount of radioactivity bound to c-cpn is lower (compared to GroEL) by a factor of about 10. Upon incubation with ATP, GroEL and c-cpn appear to discharge their bound *E. coli* proteins to a similar extent (Fig. 2b, c).

* To whom correspondence should be addressed.

Because the mechanisms of action of GroEL and c-cpn both seem to involve the unfolding and release of target proteins, one might expect that these (and other) chaperonins could function interchangeably, facilitating the folding of whichever target proteins they could bind and release. We therefore tested whether GroEL or mt-cpn could fold actin or tubulin, the major natural substrates for folding by c-cpn^{5,11-15}. Although GroEL binds labelled denatured β -actin, no product recognizable as native actin was detected upon incubation with ATP, with or without

GroES (Fig. 3a, b). Excess GroEL was included in these reactions; without this excess, there was significant loss of GroEL-bound target protein, but no generation of native product (data not shown). To probe the state of the bound target protein, an eightfold molar excess of c-cpn was included in the folding reactions. Interestingly, radioactive target protein was discharged from the GroEL binary complex, with a corresponding appearance of radioactivity associated with c-cpn and a gradual accumulation of native β -actin (Fig. 3c). Parallel

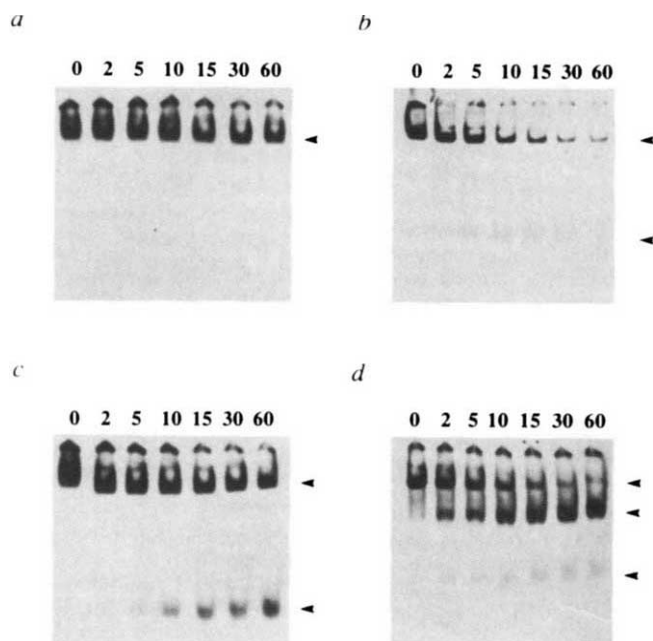
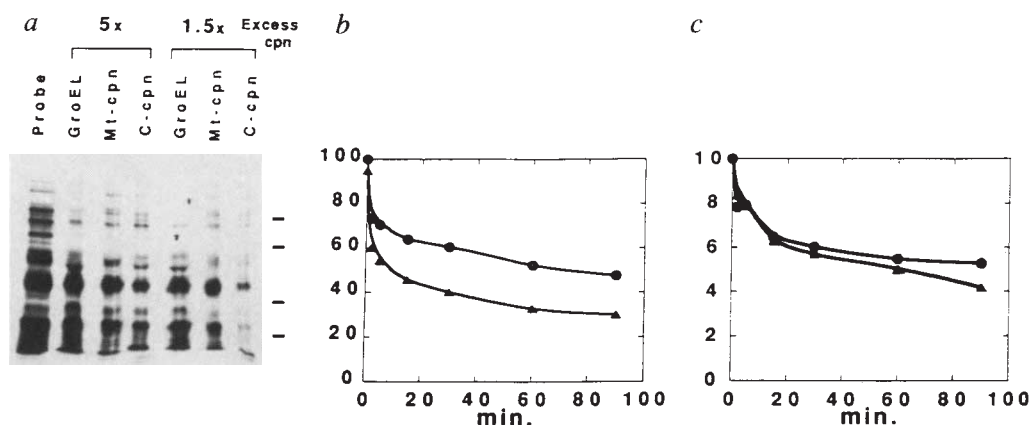


FIG. 1 Cycling of β -actin by c-cpn. a-d, Kinetic analysis on non-denaturing polyacrylamide gels of the products of *in vitro* folding reactions containing β -actin/c-cpn binary complex. Reaction products obtained following incubation of dilute β -actin/c-cpn binary complexes with ATP, either quenched with hexokinase (a) or unquenched (b-d), without added chaperonin (b), with enough added c-cpn to provide optimal reaction conditions (c), or with a 15-fold molar excess (relative to the initial β -actin/c-cpn binary complex) of mt-cpn (to act as a trap for released non-native forms¹⁶) (d). Arrows (top to bottom) show the location of β -actin/c-cpn binary complex, mt-cpn/ β -actin binary complex and native β -actin, respectively.

METHODS. The c-cpn/ β -actin binary complexes were generated by diluting 1 μ l (containing about 10 pmol) of ³⁵S-labelled urea-denatured target protein⁵ into 100- μ l reactions containing 20 mM KCl, 1 mM MgCl₂, 1 mM EGTA, 1 mM DTT, 20 mM MES, pH 6.8 (folding buffer) containing nucleotide-free c-cpn⁵ (12 pmol). The mixture was incubated for 10 min at 30 °C and applied to a 0.5 $\times$ 10 cm column of Superose 6 equilibrated and run at 4 °C in folding buffer supplemented with 0.1 M KCl. The dilute (25 nM) binary complex emerging from the Superose column (0.15 ml) was incubated with 1 mM ATP, with or without 10 mM glucose and 2.5 units of hexokinase (Boehringer), plus purified rabbit haemoglobin (5 mg ml⁻¹) as a stabilizing reagent. In some experiments, the reactions contained either an 8-fold molar excess of c-cpn or a 15-fold molar excess of mt-cpn¹⁷. At various times (shown in min), aliquots (20 μ l) were withdrawn from the reactions and flash-frozen. Reaction products were resolved by electrophoresis on 4.5% non-denaturing polyacrylamide gels^{5,11}.

FIG. 2 Recognition of unfolded *E. coli* target proteins by GroEL, mt-cpn and c-cpn. a, Analysis on an 8.5% SDS polyacrylamide gel of binary complexes formed between a 5-fold or 1.5-fold molar excess of chaperonin (GroEL, mt-cpn or c-cpn) and an unfractionated mixture of labelled unfolded *E. coli* proteins. Molecular size markers (97, 66, 45, 31K) are shown on the right. b, c, Discharge in the presence of Mg-ATP of binary complexes formed between unfractionated *E. coli* target proteins and GroEL (b) or c-cpn (c) in the absence (●) or presence (▲) of cofactors (GroES in reactions containing GroEL, cofactors A and B¹¹ in reactions containing c-cpn). Data shown is the average of two experiments, and is expressed as a percentage of radioactivity contained in input GroEL binary complex.

METHODS. A 1 ml culture of *E. coli* (strain DH5 α ; A₅₅₀ = 0.1) was labelled in minimal medium by growth at 37 °C for 4 h in the presence of 1 mCi of ³⁵S-methionine. Cells were collected and disrupted by vortexing in 50 μ l of 8 M guanidine-HCl, 10 mM DTT, 5 mM EDTA, 20 mM Tris-HCl, pH 7.4. Insoluble material was removed by centrifugation, and the supernatant applied to a column (0.5 $\times$ 10 cm) of Sephadex G25 equilibrated and run in 8 M urea, 10 mM DTT, 1 mM EGTA, 20 mM Tris-HCl, pH 7.4; the specific activity of the labelled proteins emerging from this column was 7–10 $\times$ 10⁶ c.p.m. μ g⁻¹. This probe (2.5 pmol, calculated



assuming an average M_r of 40K) was diluted into folding buffer (40 μ l) containing either 3.6 pmol or 12 pmol of GroEL, mt-cpn or c-cpn and incubated at 30 °C for 10 min. Binary complexes were isolated by fractionation on 4.5% non-denaturing polyacrylamide gels⁵. Following staining and destaining, bands containing binary complex were excised, rinsed in water, soaked for 20 min in several changes of SDS gel-loading buffer, and inserted directly into the slots of an 8.5% SDS polyacrylamide gel. In the experiments shown in b and c, discharge of binary complexes isolated by gel filtration on Superose 6 (see Fig. 1) was initiated by addition of ATP to 1 mM and incubation at 30 °C. At various times, reaction products were analysed on non-denaturing polyacrylamide gels as described⁵.

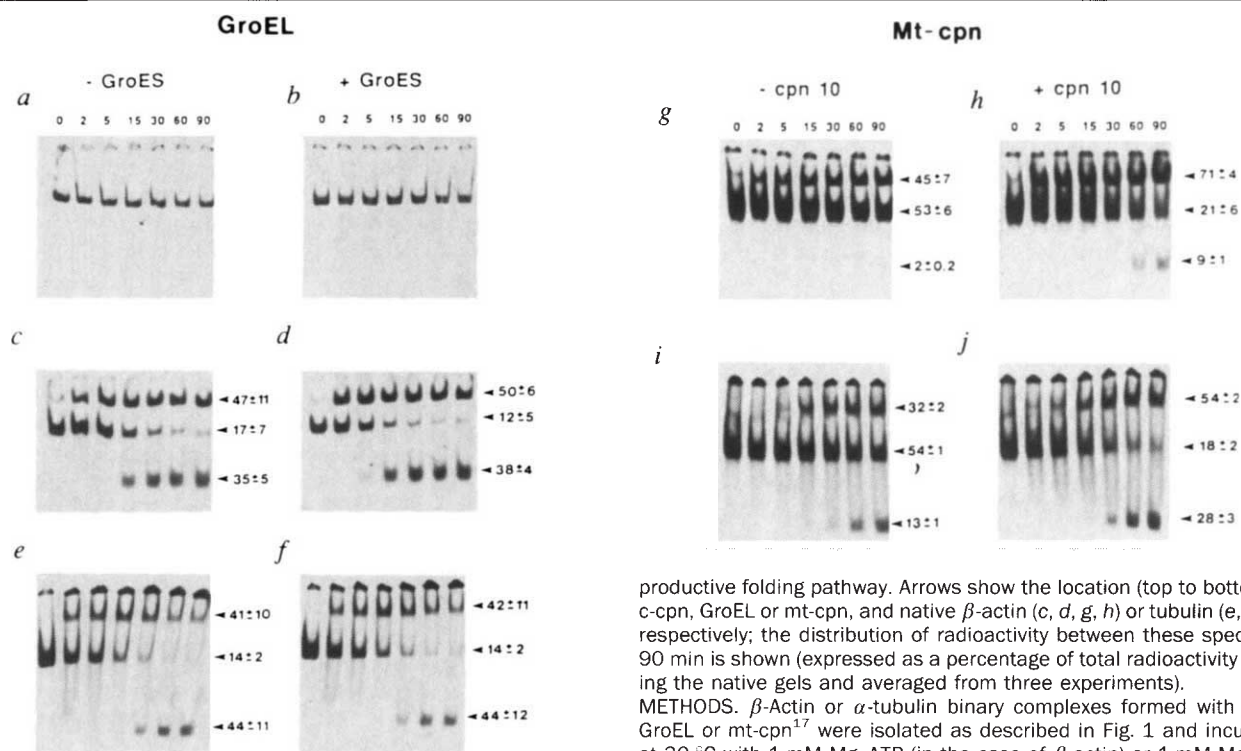


FIG. 3 GroEL and mt-cpn can bind and cycle β -actin or α -tubulin, but cannot aid their productive folding, with or without their cognate cochaperonins (GroES and cpn10). *a, b*, Unfolded β -actin forms a stable binary complex with GroEL, but is not partitioned to the native state upon incubation with ATP, with or without GroES. *c-j*, β -actin (*c, d, g, h*) or α -tubulin (*e, f, i, j*) target protein is released from GroEL (*c-f*) or mt-cpn (*g-j*) with (*d, f*) or without (*c, e*) GroES or with (*h, j*) or without (*g, i*) cpn10 in a form that can be recognized by c-cpn and partitioned to a

productive folding pathway. Arrows show the location (top to bottom) of c-cpn, GroEL or mt-cpn, and native β -actin (*c, d, g, h*) or tubulin (*e, f, i, j*), respectively; the distribution of radioactivity between these species at 90 min is shown (expressed as a percentage of total radioactivity entering the native gels and averaged from three experiments).

METHODS. β -Actin or α -tubulin binary complexes formed with either GroEL or mt-cpn¹⁷ were isolated as described in Fig. 1 and incubated at 30 °C with 1 mM Mg-ATP (in the case of β -actin) or 1 mM Mg-ATP, 1 mM Mg-GTP, cofactors A and B and carrier native tubulin¹¹ (in the case of α -tubulin). In some experiments (shown in *a* and *b*), the reactions were supplemented with an eightfold molar excess (relative to the input binary complexes) of GroEL; in others (shown in *c-j*) the folding reactions contained an eightfold molar excess of c-cpn. Reactions shown in *b, d, f* and *h, j* were also supplemented with either a 10-fold molar excess of GroES (*b, d, f*) or a 2-fold molar excess of either GroES (*d, f*) or cpn10¹⁸ (*h, j*). At various times, aliquots were withdrawn from the reactions and analysed as described^{5,11}.

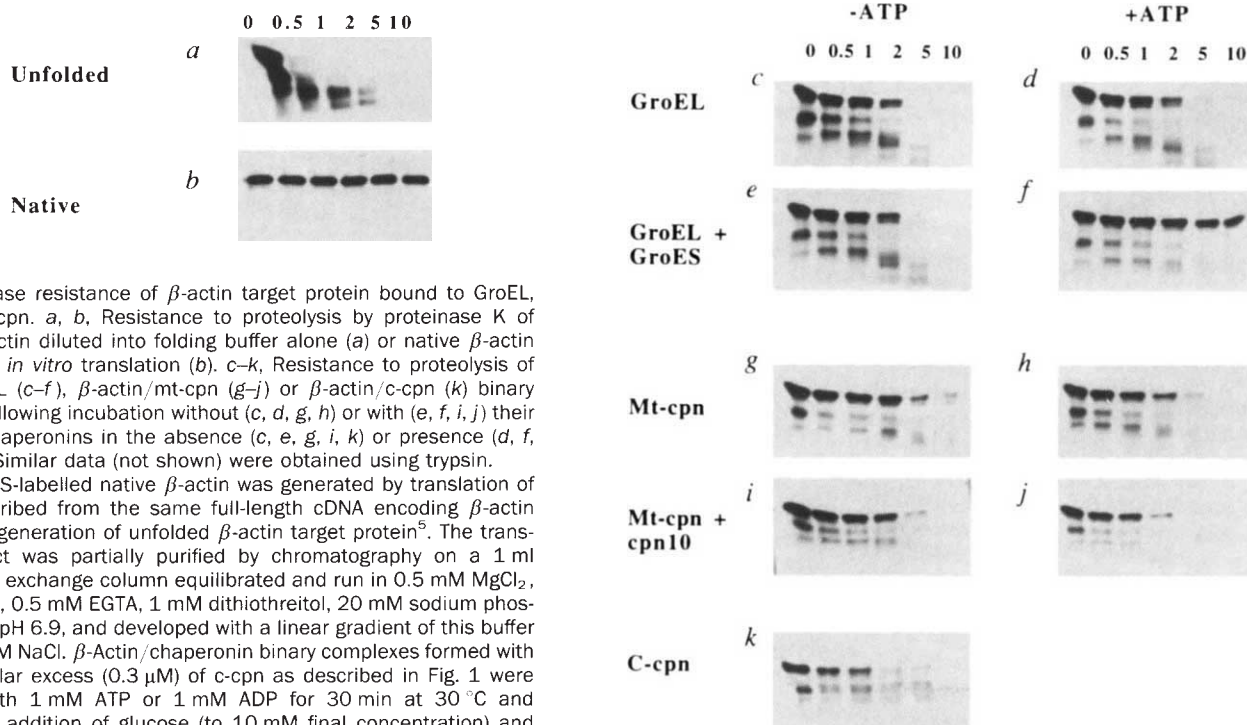


FIG. 4 Protease resistance of β -actin target protein bound to GroEL, mt-cpn or c-cpn. *a, b*, Resistance to proteolysis by proteinase K of unfolded β -actin diluted into folding buffer alone (*a*) or native β -actin generated by *in vitro* translation (*b*). *c-k*, Resistance to proteolysis of β -actin/GroEL (*c-f*), β -actin/mt-cpn (*g-j*) or β -actin/c-cpn (*k*) binary complexes following incubation without (*c, d, g, h*) or with (*e, f, i, j*) their cognate cochaperonins in the absence (*c, e, g, i, k*) or presence (*d, f, h, j*) of ATP. Similar data (not shown) were obtained using trypsin.

METHODS. 35 S-labelled native β -actin was generated by translation of mRNA transcribed from the same full-length cDNA encoding β -actin used for the generation of unfolded β -actin target protein⁵. The translation product was partially purified by chromatography on a 1 ml MonoQ anion exchange column equilibrated and run in 0.5 mM MgCl₂, 0.05 mM ATP, 0.5 mM EGTA, 1 mM dithiothreitol, 20 mM sodium phosphate buffer, pH 6.9, and developed with a linear gradient of this buffer containing 1 M NaCl. β -Actin/chaperonin binary complexes formed with a fivefold molar excess (0.3 μ M) of c-cpn as described in Fig. 1 were incubated with 1 mM ATP or 1 mM ADP for 30 min at 30 °C and quenched by addition of glucose (to 10 mM final concentration) and hexokinase (2 units) so that both reactions contained equal amounts of ADP and protein. The products of these reactions were incubated with 18.5 nM proteinase K at room temperature. At various times (shown in min), aliquots were withdrawn and the proteolytic reaction quenched

by the addition of phenylmethylsulphonyl fluoride to 5 mM. Reaction products were resolved by electrophoresis on 10% SDS-tricine polyacrylamide gels¹⁹.

experiments containing GroES yielded very similar data (Fig. 3d). Thus, β -actin released from GroEL is recognized and cycled by c-cpn at least as efficiently as β -actin diluted from denaturant. Similar data were obtained using α -tubulin as target protein (Fig. 3e,f). We conclude that GroEL can form a binary complex with labelled unfolded β -actin or α -tubulin and releases them in an ATP-dependent manner, but cannot aid their proper folding, with or without GroES.

We also found that mt-cpn, like GroEL, cannot help the productive folding of β -actin or α -tubulin (data not shown); however, mt-cpn can cycle these target proteins (although to different extents compared with GroEL) (Fig. 3g-j). We conclude that actin and tubulin target proteins can both be discharged from GroEL or mt-cpn in the presence of ATP in a spectrum of states that, though recognizable by c-cpn, cannot proceed directly to the native state. Rather, they are confined to an endless cycle of binding to and release from different GroEL or mt-cpn molecules.

To look for differences in the state of chaperonin-bound target proteins, we compared their sensitivity to two different proteases (Fig. 4). In the absence of ATP, β -actin bound to GroEL, mt-cpn or c-cpn has a similar low resistance to protease, suggesting that the three chaperonins have the same ability to unfold proteins. However, protease-resistant states do accumulate when β -actin is cycled by GroEL in the presence of GroES (Fig. 4f). The protease sensitivity of β -actin cycled by c-cpn cannot be interpreted because of the accumulation of protease-resistant native product, but when α - or β -tubulin is cycled by c-cpn, protease-resistant states are generated (G.T. *et al.*, manuscript in preparation). Thus, c-cpn and GroEL/GroES produce more native-like spectra of folding intermediates than either GroEL alone or mt-cpn, with or without cpn10.

It has been proposed that chaperonins function by unfolding kinetically trapped intermediates^{7,8}. However, they do not function in a manner that mimics unfolding by urea or guanidine, because all three chaperonins can aid the folding of target proteins (even following a single round of interaction) under conditions in which no spontaneous folding would occur upon dilution from denaturant. These considerations and the data presented here imply that a distinct set of folding intermediates is released from different chaperonins. Our data suggest that chaperonin-bound actin or tubulin target proteins are extensively unfolded; however, when cycled by GroEL/GroES, and even more so when cycled by c-cpn, significant higher order structure is acquired. Thus, GroEL, mt-cpn and c-cpn seem to differ not in their ability to unfold or cycle proteins, but in their ability to generate productive folding intermediates. □

Received 9 November 1994; accepted 14 March 1995.

1. Ellis, R. J. *Nature* **328**, 378–379 (1987).
2. Gething, M.-J. & Sambrook, J. *Nature* **355**, 33–44 (1992).
3. Horwich, A. L., Low, K. B., Fenton, W. A., Hirshfield, I. N. & Furtak, K. *Cell* **74**, 909–917 (1994).
4. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **342**, 884–889 (1989).
5. Gao, Y., Thomas, J. O., Chow, R. L., Lee, G.-H. & Cowan, N. J. *Cell* **69**, 1043–1050 (1992).
6. Braig, K. *et al.* *Nature* **371**, 578–586 (1994).
7. Todd, M. J., Viitanen, P. V. & Lorimer, G. H. *Science* **265**, 659–666 (1994).
8. Weissman, J. S., Kashi, Y., Fenton, W. L. & Horwich, A. L. *Cell* **78**, 693–702 (1994).
9. Rommelaere, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **90**, 11975–11979 (1993).
10. Kubota, H., Hynes, G., Carne, A., Ashworth, A. & Willison, K. *Curr. Biol.* **4**, 89–99 (1994).
11. Gao, Y., Vainberg, I. E., Chow, R. L. & Cowan, N. J. *Molec. cell. Biol.* **13**, 2478–2485 (1993).
12. Ursic, D. & Culbertson, M. R. *Molec. cell. Biol.* **11**, 2629–2640 (1991).
13. Yaffe, M. B. *et al.* *Nature* **358**, 245–248 (1992).
14. Chen, X., Sullivan, D. S. & Huffaker, T. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9111–9115 (1994).
15. Vinh, D. B.-N. & Drubin, D. G. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9116–9120 (1994).
16. Gao, Y. *et al.* *J. Cell Biol.* **125**, 989–996 (1994).
17. Viitanen, P. V. *et al.* *J. Biol. Chem.* **267**, 695–698 (1992).
18. Hartman, D. J., Hoogenraad, N. J., Condron, R. & Hoj, P. B. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3394–3398 (1992).
19. Schagger, H. & Von Jagow, G. *Analyt. Biochem.* **166**, 368–379 (1987).

ACKNOWLEDGEMENTS. This work was supported by a grant from the NIH (to N.J.C.).

ERRATUM

Glycogen synthase kinase-3 and dorsoventral patterning in *Xenopus* embryos

Xi He, Jean-Pierre Saint-Jeannet, James R. Woodgett, Harold E. Varmus & Igor B. Dawid

Nature **374**, 617–622 (1995)

IN the legends to Figs 2d and 3b of this Article, the shading of histogram bars was incorrectly designated. These should be: white bar, normal embryo; hatched bar, incomplete axis duplication; black bar, complete axis duplication. □

CORRECTIONS

XLas is a new type of G protein

Ralph H. Kehlenbach, Julia Matthey & Wieland B. Huttner

Nature **372**, 804–809 (1994)

WE wish to report a discrepancy in the predicted size of XLas and its apparent size on SDS-PAGE. To express XLas, we transfected PC12 cells with a composite clone assembled from clones CDM8-XL, RACE I.24 and RACE II.7 (see Methods in Fig. 2c legend) which contains the open reading frame that runs from the ATG at nucleotides 46–48 to the CTC at nucleotides 2,581–2,583 (Fig. 2c). Surprisingly, this composite clone, which encodes a polypeptide of relative molecular mass 92K, yields a protein with an apparent mobility of 124K on SDS-PAGE. Expression in PC12 and CHO cells of the clone CDM8-XL (nucleotides 380–2,956; Fig. 2c), which encodes a polypeptide of relative molecular mass 78K, yields a protein with an apparent electrophoretic mobility of 94K, identical to the endogenous XLas of PC12 cells. Thus it appears that translation of XLas starts at the ATG at nucleotides 439–441, resulting in a 715-amino-acid protein of relative molecular mass 78K (XL portion, 367 amino acids; relative molecular mass 37K) with an anomalous electrophoretic mobility. This does not affect any of the conclusions of our study, except that the N-terminal domain preceding the EPAA repeats (Fig. 2c,d) is shorter than we assumed. The significance of the open reading frame starting at the ATG at nucleotides 46–48 remains to be investigated. □

An Early Miocene anthropoid skull from the Chilean Andes

John J. Flynn, André R. Wyss, Reynaldo Charrier & Carl C. Swisher

Nature **373**, 603–607 (1995)

THE captions of parts b and c of Fig. 1 legend in this Letter were transposed as published. □

Fluorescence polarization — a new tool for cell and molecular biology

William J. Checovich, Randall E. Bolger and Thomas Burke

Fluorescence polarization (FP) equilibrium binding assays differ from other types of binding studies in one important regard: they require no steps to separate free from bound tracer and are therefore fast, simple and accurate.

FLUORESCENCE polarization (FP) is unique among methods used to analyse molecular binding because it gives a direct, nearly instantaneous measure of a tracer's bound/free ratio, even in the presence of free tracer. Most other methods require the physical separation of free and bound tracer. The theory of FP, first described in 1926 (ref. 1), is based on the observation that fluorescently labelled molecules in solution, excited with plane-polarized light, will emit light back into a fixed plane if the molecules remain stationary between excitation and emission. Molecules, however, rotate and tumble, and the planes into which light is emitted can be very different from the plane used for excitation. The polarization of a molecule is proportional to molecule's rotational relaxation time, or the time it takes to rotate through an angle of 68.5° . Rotational relaxation time is related to viscosity (η), absolute temperature (T), molecular volume (V) and the gas constant (R).

Polarization value μ rotational relaxation time is proportional to $3\eta V/RT$.

Therefore, if viscosity and temperature are held constant, polarization is directly proportional to the molecular volume. Changes in molecular volume result from the binding or dissociation of two molecules, degradation, or conformation changes.

Fluorescence polarization analysers measure the amount of light emitted into

horizontal and vertical planes by a fluorescent molecule excited with a plane of polarized light (for example, vertical plane), and then calculate a polarization value according to the following equation. The polarization value P is defined as the difference of the vertical and horizontal intensities divided by the sum of these intensities.

High polarization values mean that the molecule is large and has rotated little during the excited state; low values mean that the molecule is small and rotated very quickly during the excited state. A term sometimes seen in the FP field is 'anisotropy', which can usually be used interchangeably with 'polarization'. Anisotropy and polarization are mathematically closely related and most instruments can calculate both.

Equilibrium binding studies

Fluorescence polarization can be used to analyse most molecular interactions, including DNA-protein, protein-protein and antigen-antibody binding². Because FP assays do not require the separation of bound and free tracer and because all measurements are made in solution, true equilibrium is achieved. In addition, because FP measurements do not adulterate the samples, they can be treated and reanalysed in order to ascertain the effect on binding of changes in parameters such as pH, temperature and salt concentration. FP measurements can also be taken in real-time, allowing the kinetic analysis of association and dissociation reactions.

FPIA

More than 50 fluorescence polarization immunoassays (FPIA) are currently commercially available. These are homogeneous competitive immunoassays that differ from radioimmunoassays and ELISAs in one important regard — FPIAs require no steps to separate free and bound tracer³. In an FPIA, antibody and fluorescent antigen are com-

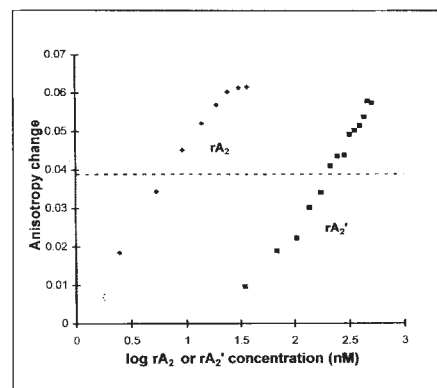


FIG. 2 Effect of thrombin activation on factor XIII A and B subunit association. Aliquots of rA_2 or thrombin-activated rA_2' were added to a starting solution of B_2 (19 nM; initial anisotropy value = 0.1865). The horizontal dotted line shows the half maximum anisotropy change.

bined such that most of the antigen is bound, thereby maximizing its polarization value. The greater the concentration of unlabelled antigen added to this system, contributed by either standards or unknowns, the greater the amount of labelled antigen that is competitively released from the antigen-antibody complex. As the fraction of unbound labelled antigen increases relative to that still bound, the polarization value drops. The sensitivity of FPIAs is affected primarily by the sensitivity of the instrument used, the affinity of labelled antigen for the antibody, and the difference in polarization values (molecular size) between the free antigen and the antibody-antigen complex. FPIAs are therefore best suited for the quantitation of small antigens, even peptides. ELISAs, on the other hand, are often difficult to develop for small molecules because multiple antigenic sites are required.

Wei and Herron recently described an FPIA for the detection of a large antigen, serum human chorionic gonadotropin (hCG; 237 amino acids)⁴. They identified a fluorescently labelled hCG 13-base peptide that reacted strongly ($K_d = 66$ nM) with an anti-hCG monoclonal antibody. The competitive calibration curve (Fig. 1) had a limit of sensitivity of 1 nM unlabelled hCG (0.46 IU ml^{-1}).

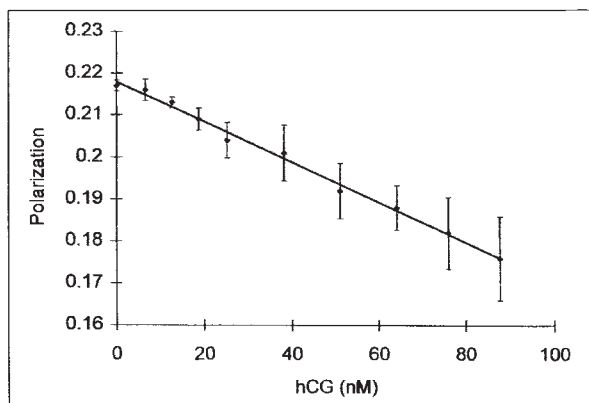


FIG. 1 Calibration curve of an FPIA for human chorionic gonadotropin (hCG). Aliquots of hCG stock solution were added to a mixture of antibody (1.5×10^{-8} M) and tracer (4×10^{-9} M). Error bars represent the standard deviation of three separate experiments.

Protein-protein interactions

Fluorescence polarization equilibrium binding studies usually involve titrating a small amount of fluorescently labelled protein with a range of concentrations of a second protein. As the fraction of bound tracer increases, the polarization value rises. Equilibrium binding constants are calculated from the resulting binding curve (for example, polarization versus unlabelled protein concentration). If polarization changes are followed over time, kinetic parameters (for example, k_1 and k_{-1}) can be determined.

Radek *et al.* have used FP to study the activation of blood coagulation factor XIII, a tetramer composed of two homodimers, A_2 and B_2 (ref. 5). Thrombin proteolyzes A_2 to A_2' , leading to the calcium-dependent dissociation of B_2 from A_2' . Radek and his colleagues added increasing concentrations of recombinant A_2 (rA_2) or thrombin-activated A_2' to fluorescently labelled native B_2 and measured anisotropy changes (Fig. 2). Analysis of the binding isotherms indicated that thrombin cleavage of rA_2 to rA_2' results in a dramatic decrease in the affinity of rA_2 for labelled B_2 (294 nM for rA_2' versus 19 nM for rA_2), thereby explaining A_2B_2 complex dissociation.

DNA-protein interactions

Fluorescence polarization analysis of DNA-protein binding is identical to the protein-protein studies described above except that a fluorescently labelled oligonucleotide is used instead of a peptide. LeTilly and Royer were the first to use FP to directly calculate dissociation constants for a specific protein-DNA interaction: single and tandem dimer *trp* repressor protein (TR) complex formation with *trp* operator DNA (ref. 6). Figure 3 shows binding isotherms for the

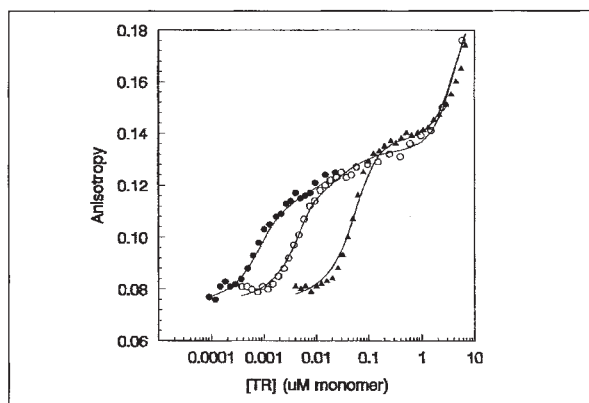


FIG. 3 Binding isotherm of fluorescein labelled *trp* operator DNA and *trp* repressor (TR) protein. Various concentrations of TR were titrated with 300 pM (●), 1.5 nM (○) or 30 nM (▲) of a fluorescein-labelled 25-base oligonucleotide containing the *trp* operator sequence. The curves represent a least-squares nonlinear regression curve fit of the data. The final, steep rise is probably due to nonspecific binding.

interactions of a labelled 25-base oligomer containing the *trp* operator sequence with a range of concentrations of TR (p ; [DNA], K_d). These authors ascribed the initial increase in polarization to the binding of the first TR dimer ($K_d = 0.13$ nM), and the second, more gradual rise to binding of the second TR dimer ($K_d = 12.5$ nM). This difference in affinities for the first and second TR dimer binding was attributed to an imperfection in the second half-site binding domain on the operator DNA.

Degradative assays

The action of degradative enzymes, such as proteases, DNases and RNases, results in a change in molecular size, and can therefore be followed with FP. Fluorescence polarization assays for protease^{7,8}, DNase⁹ and RNase¹⁰ involve the degradation of a large, intact, fluorescently labelled substrate (for example, F-casein, F-DNA or F-RNA having a high polarization value) to smaller, low-polarization fragments. Assays can be formatted for real-time kinetic analysis or for end-point determination.

Figure 4 shows a time course of trypsin activity using fluoresceinated casein⁷. Almost any substrate can be fluorescently labelled and used to measure the activity of a specific enzyme (for example, for HIV protease, tissue metalloproteases). The standard curve for the casein-based assay was linear over at least a 100-fold concentration range, and the detection limit was 10 times more sensitive than was possible with more common acid-precipitation assays (Ref. 7). Only radioactive protease assays offer greater sensitivity than FP assays.

High-throughput screening

The most recent advance in FP technology is the introduction of a 96-well fluorescence polarization plate reader (Jolley Consulting and Research, Round Lake, Illinois). DNA-protein, protein-protein and enzyme FP assays have all been converted to the microtitre well format and are particularly well suited for drug discovery, where thousands of candidate compounds must be screened. Microtitre-well FP assays offer a tremendous reduction in time compared with older methods that require precipitations, transfers or filter binding.

Practical considerations

There are three primary concerns when designing

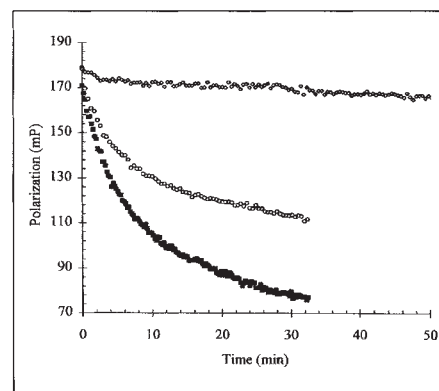


FIG. 4 Time course of fluorescein-labelled casein digested with trypsin. The reaction was started by addition of 5 pmol fluorescein-labelled casein. Fluorescence polarization was measured every 13 s on the Beacon Fluorescence Polarization System (PanVera Corporation., Madison, Wisconsin). ■, 200 ng trypsin; ○, 100 ng trypsin; ◇, no trypsin added. mP, milli-polarization units.

FP experiments: (1) molecular weight of the fluorescently labelled molecule, (2) adventitious fluorescence in the buffers and biological samples to be tested, and (3) the sensitivity of the fluorescence polarization detection instrument. Because very large molecules tend to rotate and tumble very little during the fluorescence lifetime, there is a practical upper molecular weight limit for the fluorescently labelled molecule of about 100,000 daltons, depending on the local mobility of the labelled domains. Therefore, whenever possible, the smaller of two interacting molecules should be labelled in order to maximize polarization shifts.

As the sensitivity requirement for an FP experiment increases, complications caused by background fluorescence in buffers and biological samples also increase. It is prudent to use chemicals certified as having a low fluorescence background. Fortunately, fluorescence polarization values are not usually influenced by changes in intensity. Therefore, FP can be used with relatively opaque or turbid samples that cannot be analysed with other optical methods.

Finally, as with most technologies, the sensitivity of FP can be limited by the quality of the instrumentation. Most fluorescence detectors retrofitted for measuring polarization have difficulty keeping light polarized as it passes through long distances and across numerous mirrors and monochrometers. These *ad hoc* configurations result in sensitivities in polarization mode usually no greater than the nanomolar range. Instruments dedicated to FP not only have sensitivities in the low picomolar range, but are far less expensive than converted fluorimeters. Two criteria should be met by any sensitive fluorescence polarization analyser: dilution of

fluorescence intensity is linear down to the femtomolar range, and polarization values are independent of fluorescence intensity down to the low picomolar range. We chose to use the Beacon Fluorescence Polarization System in our studies of a variety of protein-protein and DNA-protein interactions with K_d values in the low nanomolar range because it met these criteria. □

William J. Checovich, Randall E. Bolger and Thomas Burke are at the PanVera Corporation, Madison, Wisconsin 53711, USA. For more information fill in reader service number 100.

1. Perrin, F. J. *Phys. Radium* **7**, 390-401 (1926).
2. Dandliker, W. B., Hsu, M. L., Levin, J. & Rao, B. R. *Meth. Enzym.* **74**, 3-28 (1981).
3. Jolley, M. E. *J. analyt. Tox.* **5**, 236-240 (1981).
4. Wei, A. P. & Herron, J. N. *Analyt. Chem.* **65**, 3372-3377 (1993).

5. Radek, J. T., Jeong, J. M., Wilson, J. & Lorand, L. *Biochemistry*, **32**, 3527-3534 (1993).
6. LeTilly, V. & Royer, C. A. *Biochemistry* **32**, 7753-7758 (1993).
7. Bolger, R. & Checovich, W. *BioTechniques* **17**, 585-589 (1994).
8. Maeda, H. *Analyt. Biochem.* **92**, 222-227 (1979).
9. Yonemura, K. & Maeda, H. *J. Biochem.* **92**, 1297-1303 (1982).
10. Bolger, R. & Thompson, D. *Am. biotech. Lab.* **12**, 113-116 (1994).

Molecular mania

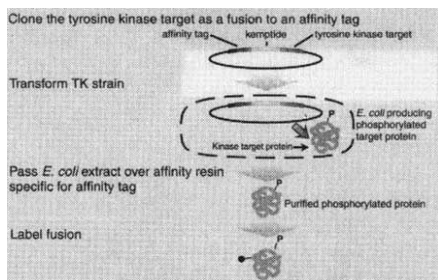
From Berlin to San Francisco researchers can find an abundance of new products such as DNA sequencers, amplification reagents, an electroporator and a checkerboard hybridization system.

IN ADDITION to Lysyl endopeptidase (Lys-C), Wako offers two new **sequencing grade enzymes** — endoproteinase Asp-N and Glu-C (*Reader Service No. 101*). Endoproteinase Asp-N, isolated from a *Pseudomonas fragi* mutant strain, specifically cuts the aspartate amino end of residual radicals of both proteins and peptides. Endoproteinase Glu-C, isolated from *Staphylococcus aureus* V8, specifically cuts the C-end of glutamic acid when using bicarbonate ammonium buffer solution (pH 7.8) or ammonium acetate buffer solution (pH 4.0) and specifically cuts the carboxy end of glutamic acid and aspartic acid when using phosphoric acid buffer solution (pH 7.8). In addition, Wako has added a new size for Lysyl endopeptidase, 2AU, which should be more convenient for the small laboratory or first-time users.

The new checkerboard **hybridization system** from Immunetics is designed to handle simultaneous hybridization of up to 1,350 probes on a single membrane (*Reader Service No. 102*). Target DNA sequences, either purified or in cell lysates, are first applied to the membrane in horizontal lines using a MiniSlot vacuum aspiration device. Membranes are then transferred to a Miniblotter where probes are introduced into parallel vertical incubation channels. Hybridizations may be carried out according to standard protocols using chemiluminescent or radiolabelled probes. The resulting checkerboard hybridization can be quantified by scanning densitometry.

■ Amplification

TaqPlus DNA polymerase is an optimized mixture of Stratagene's TaqExtender PCR additive and Taq DNA polymerase (*Reader Service No. 103*). The addition of the TaqExtender PCR additive reduces the mismatch pausing associated with Taq DNA polymerase and significantly enhances the reliability, yield and length of PCR amplified products. The



Stratagene TaqPlus DNA polymerase.

manufacturer claims that the primer-template combinations covering a wide variety of targets up to 35 kb in length can be obtained by using TaqPlus. Such long-range PCR reactions can have a broad impact on the mapping and sequencing of human genomic and cDNA sequences, the PCR-based characterization of carriers of clinically relevant large insertions or deletions, the amplification of whole virus genomes and the like.

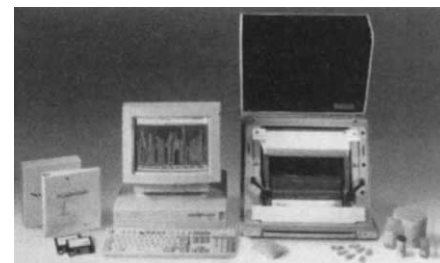
Invitrogen has introduced HotWax beads, which are designed to make 'hot start' PCR reactions easier, faster and more efficient (*Reader Service No. 104*). The beads release $MgCl_2$ into the reaction when the wax bead melts during the first denaturation step. This system is designed to eliminate the excessive manipulation required for a manual 'hot start'. In addition, the melted bead forms an evaporation barrier, making an oil overlay unnecessary. The beads are available in low (1.5 mM), medium (2.5 mM) and high (3.5 mM) $MgCl_2$ concentrations and are provided with one of four Mg^{2+} -free 5× PCR buffers (pH 8.5, 9.0, 9.5 or 10.0).

OmniStrips from Hybaid are **thermal-cycling strips** composed of eight 0.3-ml OmniTubes (*Reader Service No. 105*). These tube strips have been designed for use with the 96-well OmniBlock — the heating block available with the

OmniGene and Omn-E thermal cyclers. The profile of the tube matches that of the block wells to ensure maximum surface contact. In addition, thin-walled polypropylene construction provides thermal conduction to the sample. Tubes can be supplied as a standard 8 × 12 array, pre-racked into plates of 96 tubes. Alternatively, strips are available in individual strips of eight. A 96-position rack is also available, providing a convenient platform for reagent dispensing. Separate caps are supplied, also in strips of eight.

■ Sequencing

Pharmacia Biotech has introduced ALFexpress, a new automated DNA sequencer that is designed to combine high speed, accuracy and affordability for sequencing and fragment analysis (*Reader Service No. 106*). The system includes a complete range of gel casting chemicals and sequencing reagents, including the AutoLoad solid phase kit, to facilitate sample preparation and improve consistency and reliability; a self-contained sequencing unit, with few moving parts for quiet and reliable operation; and a standard PC-based computer that runs a multitask operating system, with software that provides a high-resolution picture of the sample analysis. The system uses single-dye chemistry to eliminate spectral overlap and filtering problems and improve the signal-to-noise ratio. ALFexpress is also equipped with a high-performance base-calling algo-



Pharmacia's ALFexpress DNA sequencer.

rhythm that is capable of reading 300 base pairs per hour, with a stated accuracy of 99 per cent. It is said to do fragment analysis up to 150 to 200 base pairs in about 30 min.

Molecular Dynamics offers a new 8-page, full-colour brochure for the **Vistra DNA sequencer 725** (*Reader Service No. 107*). The system is comprised of instrumentation software and reagents and is designed to replace manual, film-based



ABI DNA sequencer from Perkin-Elmer.

sequencing methods. The system has automated gel electrophoresis, band detection, lane tracking and sequence assignment. Its features are depicted in the brochure with photographs, illustrations and screen-capture graphics. A special section lists the sequencing reagents available for use with the system.

The new **ABI Prism 377 DNA sequencer** is now available from Perkin-Elmer (*Reader Service No. 108*). The system uses new fluorescent detection technologies along with a patented multicolour labelling system and improved software to achieve a rate of sequencing that is said to be four times faster. The manufacturer states that the new sequencer can collect data at rates of up to 7,200 bases per hour — compared with 1,800 bases per hour available from its predecessor.

A new **DNA sequencing catalogue** is now available from Amersham Life Science (*Reader Service No. 109*). It contains details of three new Sequenase kits: the Sequenase PCR product sequencing kit, the DTAq cycle sequencing kit and the Plasmid Quick-Denature sequencing kit. In addition, application guides have been included to show how sequence compression artefacts can be identified and corrected, how manganese ions can help to read sequences close to the primer, and how pyrophosphatase can be used to prevent DNA polymerase from working in reverse. This comprehensive catalogue contains details of all Amersham's sequencing products that cover the entire sequencing process. These products include modifying and restriction

enzymes, radioactive and nonradioactive nucleotides, gel products and autoradiography accessories.

The Kodak IBI Base Runner **nucleic acid sequencers** are designed to provide faster set-up, more reproducible results and greater flexibility than conventional sequencers (*Reader Service No. 110*). There are two models available from Sigma Aldrich Techware — the Base Runner 200 features a dual-gel design for maximum sample capacity and the Base Runner 100 is an economical, single-gel sequencer. Both models have height-adjustable gel assembly to accommodate gels from 30 to 60 cm in length and are designed to be capable of resolving 400 or more bases per run. The systems are designed to eliminate gel artefacts and minimize pre-run time by using Thermocore plates. The removable tank has drains to allow safe and convenient disposal. Each unit is shipped complete with 60-cm plates, combs, spacers and the other accessories required for nucleic acid sequencing.

■ Separation and concentration

The new λ Quick Spins kit from Stratatech Scientific has been refined to **isolate λ DNA** from agarose plate lysates (*Reader Service No. 111*). This kit is designed to provide high-quality DNA that is ready to cut, sub-clone, sequence or amplify. All items are provided to isolate pure λ DNA starting with the growth media and all the reagents necessary to isolate and purify phage DNA.

The DNA Mini system is a compact and reliable complete **vacuum concentration system** dedicated to drying and concentrating RNA/DNA precipitates (*Reader Service No. 112*). This system is designed to be easy to use, environmentally friendly and compact. It consists of a VR-mini centrifuge, a Teflon membrane pump, a rotor and a built-in analogue vacuum read-out and autovent.

Savant's Gene Transformer GTF100 **electroporator** features a compact design with built-in power supply and cuvette holder (*Reader Service No. 113*). The GTF100 is suited for cell transformation using standard electroporation protocols. This system has been designed to be easy to operate, with three optimized voltage settings, and to provide consistent transformation efficiencies for both bacterial and mammalian cells. It does not require a separate power supply, making it portable and easy to situate in the laboratory. A safety switch prevents power discharge when the cover is open.

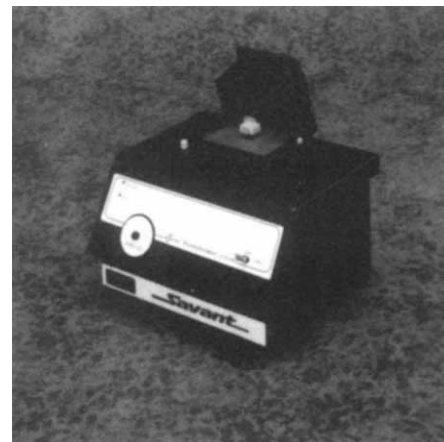
New from Labconco is the CentriVap DNA system, the latest addition to the company's line of **concentrators and cold**

traps (*Reader Service No. 114*). The system is designed to process DNA samples or other minute quantities of solvents. It includes a benchtop concentrator, DNA rotor, explosion-proof DNA vacuum pump, an accessory inlet glass trap and all connecting tubing. CentriVap uses centrifugal force along with vacuum to evaporate solvents from multiple small samples. There is a new built-in vacuum gauge for easy monitoring and a new automatic bleed valve that releases vacuum any time the rotor stops. It has five temperature settings (ambient, 35 °C, 45 °C, 60 °C and 75 °C) eliminating the need for an infrared heater. The rotor holds 48 0.5-ml or 36 2.0-ml sample tubes.

The *Tth* **thermostable DNA polymerase** has both DNA-dependent and RNA-dependent polymerase activities (*Reader Service No. 115*). Isolated from *Thermus thermophilus*, *Tth* DNA polymerase has a half life of over one hour at 95 °C and optimal activity at temperatures over 65 °C. These properties allow for DNA synthesis from either RNA or DNA templates that may be recalcitrant to DNA synthesis reactions at lower temperatures due to the formation of template secondary structures. The product is supplied with a 20× reaction buffer and separate tubes of MgCl₂ and MnSO₄ to allow for the optimization of divalent cation levels with any given template.

■ Transformation

Savant's Gene Transformer GTF100 **electroporator** features a compact design with built-in power supply and cuvette holder (*Reader Service No. 116*). The GTF100 is suited for cell transformation using standard electroporation protocols. This system has been designed to be easy to operate, with three optimized voltage settings, and to provide consistent transformation efficiencies for both bacterial and mammalian cells. It does not require a separate power supply, making it portable and easy to situate in the laboratory. A safety switch prevents power discharge when the cover is open.



Gene Transformer from Savant.

PRODUCT REVIEW

The Tfx-50 reagent from Promega is a cationic lipid used for **transfecting eukaryotic cells** (Reader Service No. 117). It is supplied as a dried lipid film that forms multi-lamellar vesicles that associate with nucleic acids upon resolubilization. These vesicles facilitate the transfer of nucleic acids into cells by fusion with the cell membrane. Tfx-50 can be used for transfection of a number of cell lines including HeLa, HepG2, NIH/3T3 and COS cells. This reagent allows the use of serum for cell lines that are sensitive to the absence of serum. It comes packaged in three separate 0.7-mg vials.

Trevigen has announced the development of **chemically competent PMC103 host cells** for the cloning of plasmids containing genes classified as 'difficult to clone' (Reader Service No. 118). The new host cells have been shown to maintain DNA with secondary structures, methylated DNA and DNA that is A/T rich. The manufacturer states that the PMC103 transformation efficiency is a minimum of 10^9 transformants per mg of input pUC18 plasmid DNA. This is stated to be about 10 times higher than that previously attained using only the electrocompetent format.

■ Miscellaneous

Savant's Gene Transformer GTF100 **electroporator** features a compact design with built-in power supply and cuvette holder (Reader Service No. 119). The GTF100 is suited for cell transformation using standard electroporation protocols. This system has been designed to be easy to operate, with three optimized voltage settings, and to provide consistent transformation efficiencies for both bacterial and mammalian cells. It does not require a separate power supply, making it portable and easy to situate in the laboratory. A safety switch prevents power discharge when the cover is open.

Labsystem has launched a new high-throughput instrument for **simultaneous washing of 96-well microplates** (Reader Service No. 120). The instrument allows the washing process to be viewed continuously. The user may choose the parameters needed to load a microplate and simply press the 'start' button to initiate the programme. The system is designed to provide easy access to needles and features a maintenance rinse to prevent clogged needles. It has a tubular



Labsystems' Multiwash 96.

manifold design, vacuum monitoring for process reliability and colour-coded fittings for easy installation. The volume resolution for the system is 50 μ l with a volume of 300 μ l volume accuracy is said to be ± 10 per cent from well to well. The wash time for a three-cycle plate wash is 35 s for 300 μ l, with a residual volume of ± 5 μ l. □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

DNA SEQUENCING

Using an ABI 373
Automated Sequencing System

£45 per run

- PCR products
- M13 inserts
- plasmid inserts
- single stranded DNA
- double stranded DNA
- 300-400 bases of sequence

Please contact us for more information

Molecular Medicine Unit
King's College School of Medicine & Dentistry
Tel: 071- 346 3126 Fax: 071- 733 3877

Reader Service No.11

DNA Sequencing Service

for plasmids, cosmids, PCR products

one primer reaction: **DM 42,00**
both strands : * **bp DM 2,50**

*publication quality, including template purification, primer synthesis, subcloning, computer analysis

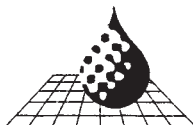
experience in automated large scale DNA sequencing for many years

MediGene GmbH

Lochhamer Str. 11, D-82152 Martinsried
Tel: 49 89 895 632 0 Fax: 49 89 895 632 20

Reader Service No.6

ADVERTISEMENT



Severn Biotech Ltd

Severn Biotech Ltd, Unit 2, Park Lane,
Stourport Road, Kidderminster, Worcs DY11 6TJ, U.K.

Custom DNA & RNA Oligo's

Probes, Primers, Antisense Oligo's, Ribonucleotides, 5' and 3' labelling with Biotin Fluorescein, Rhodamine, Modified bases purification etc.

Complete Peptide Service

Small Scale Multiple Peptides 5-10mgs or larger scale 1000mgs using L and D Amino Acids, screening grade of HPLC purified, Phosphorylated, Acetylated Amidated and Protein Carrier Conjugated (KLH, PPD, etc.)

Customer Services

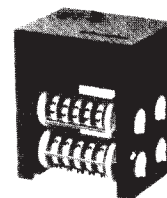
Comprehensive Custom Service for DNA, RNA and Peptides from SBL

- DNA Synthesis
- RNA Synthesis
- Multi Peptide Synthesis (5-10mgs)
- Larger Scale Peptide Synthesis 1000mgs
- HPLC Purification
- Custom Conjugation to Carriers
- Peptide Sequencing / Mass Spec.

For a Fast Efficient Service Reply Today
Telephone: (01562) 825286 Fax: (01562) 825284

Reader Service No. 80

DIABERM® EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

DIABERM
Gerate
P.O. Box 650126,
D-81215 Munich

Reader Service No.12

Future brightens for genomics in Germany

Munich. Despite the strength of its basic science, Germany got off to a late start in genetics-related research, largely as a result of public antagonism stemming from Germany's fear of its Nazi past. But a concerted campaign to influence public opinion, backed by achievements in medical genetics, has recently begun to improve the general situation.

The ministry of research (Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie, or BMBF), as part of its campaign to develop advanced science in Germany, is currently completing the first study designed to determine the precise employment situation in biotechnology.

This study covers all areas of biotechnology — including medicine, nutrition, agriculture and plant genetics, environmental science, veterinary medicine — in both public and private sectors. It reveals that there has recently been a significant growth in employment, although this is expected to level off in 1998.

In particular, the study has found that about one third of all biotechnology researchers in industry employ genetic engineering. Conversely, those industries that pay special attention to genetic engineering are heavily research-oriented — and are the fastest growing.

Recent successes in medical genetics have helped to counter a level of deep-seated public opposition to genetics research not encountered in other countries. As a result, the prospects for German molecular geneticists are at last beginning to look up.

The detailed results of this survey will be published later this year. But some of the main trends have already been made known. Two months ago, for example, Heinrich Kolb, the parliamentary state secretary for economics, reported that the number of people employed in biotechnology in Germany had doubled in the past decade, to a total of almost 40,000. This figure includes scientists, technicians and others employed in industry, universities, and research centres.

Basic research in molecular genetics is now quite strong. It already receives considerable funding from the BMBF and the Deutsche Forschungsgemeinschaft (DFG), Germany's main grant giving body, and this seems likely to rise. Most of the funding supports medical and pharmaceutical research.

In addition, Germany is at last joining international efforts to decode the human genome. Full details will be announced

next month. But the main aim of a new, eight-year programme, expected to begin later this year, will be to identify and sequence medically relevant genes, and determine their function. Jürgen Rüttgers, the research minister, has emphasized that Germany's participation in the human genome project should provide direct benefit to medicine.

The BMBF will provide DM50 million (US\$70 million) per year to support the programme, making it the largest programme in the ministry's department of biology and medicine. An international advisory committee is due to meet this week to make recommendations for the project. Although no numbers have yet been specified, the programme is certain to create a considerable number of new jobs for molecular geneticists.

Boost to gene centres

The human genome programme is expected to boost the funding of many research centres, including the four 'gene centres' that were founded by the research ministry in the early 1980s, when genetics in Germany was perceived as weak. Funding for these centres, which were intended to provide young molecular geneticists with top-class facilities and academic independence, runs out this year.

The centre in Cologne has already been dissolved, and that in Berlin will dissolve at the end of the year. Many of their scientists have had to find jobs abroad or switch to other fields. But the other two centres, in Heidelberg and Munich, have obtained further funding from their host *Länder* and other sources, and have expanded. Ernst Winnaker, director of the Munich Genzentrum, says that the junior faculty from his centre are always very successful in finding permanent positions.

Despite an environment that is becoming more welcoming for medical genetics, public acceptance of agricultural biotechnology remains low. Indeed, the two gene centres that are being dissolved, all focus on plant genetics. So too does the Institut für Genbiologische Forschung in Berlin, which was founded in 1987, currently employs about 100 scientists, but will close at the end of next year, having tried unsuccessfully to obtain funding from the government of Berlin for the period after BMBF support runs out.

Nonetheless, there are about 100 small businesses involved in plant genetics in Germany. Of these, 70 have breeding programmes, and 30 specialize in marketing foreign and German products. Together, these companies employ around 300 scientists in research and development.

But the future of genetic plant breeding

Job market remains tight for many

Ralf Schnabel, a developmental geneticist, shares the same problem that many of his fellow scientists face at present: how to find a permanent job in an extremely tight market.

Schnabel, who is 40, hopes to be appointed a full professor within a couple of years. He has all the right qualifications; having completed his PhD in Munich, three years postdoctoral research in Cambridge, England, and four years leading a small group at the Max Planck Institute for developmental biology in Tübingen, he has been an independent group leader in the Max Planck Society's (MPG's) highly competitive programme for young scientists since 1991.



Schnabel: 'trying to remain optimistic'.

This programme, which has supported nearly one hundred scientists in the life sciences since it began about twenty-five years ago, provides young investigators with independence and the chance to build up their own research group, both opportunities that in Germany are usually possible only later in one's career.

Most graduates of MPG's programme for young investigators have gone on to further academic success. Indeed a few years ago, it was not uncommon to become a director at a Max Planck Institute or a university professor. But times are getting tighter.

Although he has been successful in his research career, Schnabel says he now feels under pressure to accept any job he can get. He is looking not only in Germany, but also in central Europe. If that fails, he says he will look further afield. Although there are few academic positions available, Schnabel says that he is "trying to remain optimistic".

Toni Feder

in Germany looks less than promising. "There are too many highly qualified people looking for jobs", says Wiltrud Jennissen, a spokeswoman for the Bundesverband Deutscher Pflanzenzüchter (German Association of Plant Breeders).

Plant geneticists often leave the country to find jobs abroad. Four of six scientists who took part in a programme to promote junior faculty at the Max Delbrück Laboratory in Cologne, for example, left Germany to find academic jobs in the United States and the United Kingdom, and the programme has now been reduced by half.

This is typical for the field. "The situation is getting worse, and it is becoming more and more difficult to find jobs", says Josef Schell, a director at the Max Planck Institute for Plant Breeding in Cologne, and director of the recently dissolved gene centre there.

Attitudes vary between Germany's sixteen *Länder*. The wealthy south German *Land* Bavaria has historically been more positive about genetic research than many of the others. Boehringer Mannheim, for example, one of the largest companies involved in genetic research, has its biotechnology laboratories there, and employs over 300 scientists with doctorates.

By contrast, the state of Hessen, where Hoechst is based, has proved the most hostile. But with the general trend of increased public acceptance of genetics research, the differences between the *Länder* are fading. Several *Länder* are supporting programmes to promote new companies. In Nordrhein-Westfalen, for example, BioGenTec NRW, a firm founded last year, acts as a network between entrepreneurs, scientists, and government, and is now helping to set up ten new companies.

Confirmation that conditions for genetics-based research and development in Germany are improving has come from a recent report by the international consultancy firm Ernst and Young. A study on biotechnology in Europe, published this year, concludes that "one of the most significant and encouraging changes taking place in Europe is the new-found German enthusiasm for biotech". Specifically citing the relaxation of the country's genetic engineering laws, the report claims that Germany "could possibly become a significant target for foreign companies hoping to take advantage of its excellent science base".

The shift in public sentiment has played a major role in this process. "The atmosphere is even more important than the law itself," says Peter Stadler, head of one of Bayer AG's pharmaceutical divisions. "Now the company's decisions for expansion can be based on technical know-how rather than on the political atmosphere," he adds.

Toni Feder

France rides high on the crest of the genetics research wave

Paris. Employment prospects for geneticists in France — as in the rest of Europe — are tightly linked both to government support for life sciences and biomedical research, and to the growth of the biotechnology industry.

A key element in the latter is increasingly the emerging genome-based companies. These can be divided into three main groups: major pharmaceutical companies, who are both beginning to develop gene therapy techniques and to use genome research to revolutionize classical drug discovery; large instrument makers, who are developing new gene-based diagnostics; and a host of new companies specializing in such areas as bioinformatics, gene therapy or instrumentation.

In turn, the prospects of such industries depend strongly on two factors. One is success in genome research itself, and the other, success in its application to specific therapeutic and diagnostic goals. In each case, any successes will be linked to funding levels; support for the former comes mainly from the public purse and medical charities, while the latter is mainly supported by industry.

The growth of private industry's interest is shown by the fact that public spending on biomedical research in France now amounts to around FF9 billion (US\$1.8 billion) — or around 40 per cent of the national effort, compared with almost half in 1986. In contrast, total private funding comes to around FF14 billion, giving total research expenditure of around FF23 billion.

Life sciences research at the country's largest basic research organization, the Centre National de la Recherche Scientifique (CNRS) has, however, been affected by the wider funding crisis at the agency (see *Nature* 374, 666; 1995). The CNRS's budget for life sciences, which currently totals FF346.4 million (excluding salaries), has fallen by almost one-fifth over the past two years.

State support for genome research itself amounts to around FF80 million a year. In addition, the Généthon laboratory has been funded by the French Muscular Dystrophy Association (AFM) by FF300 million over three years.

State funding was previously funnelled through the national genome research agency (GREG). But the government has now taken direct control of the genome programme (see *Nature* 373, 650; 1995), as part of a wider plan to develop a national strategy for life sciences. This strategy is being supported by a cash injection of FF250 million over three years, which should alleviate the CNRS's diffi-

Researchers in France are among the world leaders in gene mapping and sequencing, and its industries are already deeply engaged in exploiting the results emerging from these fields. But massive investment may be needed to conserve this position.

culties in funding life sciences.

Although French spending on genome research is only around one-tenth that spent by the United States, the country has emerged among the world leaders in its efforts to map and sequence the human genome, with a run of scientific successes that includes 'second generation' physical and genetic maps.

Ironically from an employment point of view, much of this early success has stemmed from the use of robots rather than researchers by the Généthon laboratory, which was created near Paris in 1990 by the Centre d'Etudes du Polymorphisme Humain (CEPH) and AFM.

In the long-run, however, it is expected that success in such research will ultimately open up job opportunities in more intellectually satisfying areas of genetics. Aspiring geneticists might do well to heed the criticism of Daniel Cohen, CEPH's director, that "biologists spend too much time doing monkey work, leaving them less time for thinking".

Call for independence

But a report released last month by AFM argues that France risks losing the lead it has established, and being squeezed out of the next phase of the human genome project by research teams in the United States and the United Kingdom. To counter this, the report — which was commissioned from the Paris-based consultants BIPE — recommends that France consider a massive investment of between FF1.5 billion and FF6 billion in genome sequencing, in a project employing over 1000 staff and 200 to 600 sequencing machines.

The plan, which would be a joint venture between public research organizations, industry and medical charities, could create as many as 9,000 jobs in the medium-term, estimates the report. It argues that such an investment is needed to guarantee France's "national independence" in the longer term, in particular by protecting the competitiveness of its biotechnology and drug industries.

According to Bernard Barateau, the president of AFM, the sequencing of the genome "will either be done at the industrial scale or not at all." He argues that

France needs an "imaginative and strategic" political vision that is lacking at the present time. France must have a strong domestic programme, he says, if it is to become a credible partner in a proposed international collaboration to sequence the entire genome (see *Nature* 375, 93; 1995).

"We can sequence the genome in three years", predicts Barateau, who also argues that such a large-scale approach will ultimately work out cheaper than the current strategy of hunting for individual disease genes. "There isn't enough money on the face of the Earth to find all human disease genes using the current approach."

For their part, French companies have begun to multiply their initiatives to exploit genome research. In particular, the French/US drug conglomerate Rhône-Poulenc Rorer has created an ambitious international network of academic research groups and biotechnology companies to speed up the commercialization of gene therapy (see *Nature* 372, 210; 1994).

This move coincides with a decision by the ministries of health and research to coordinate their support for genome research and its clinical applications under the umbrella of a single so-called 'Génome-Santé' programme.

Declan Butler

A guide to biotechnology training courses in France, Guide des Formations 'Bio', du bac à la Thèse, is available from Biofutur, Editions Scientifiques Elsevier, 141 rue de Javel 75747 PARIS CEDEX 15.

Europe faces lack of skilled recruits

Employment in genetics-related industries is expected to grow rapidly across the European Union. Will the production of suitably qualified graduates keep up?

Paris. The growth of biotechnology industries will create 2 million jobs in the 15 member states of the European Union by the year 2000, according to the European Commission. The food, chemistry and pharmaceutical sectors, it claims, already account for almost a fifth of all European jobs — a total of 15 million positions.

This optimistic outlook has been confirmed by a recent survey of European biotechnology, carried out by the management consultants Ernst and Young. The survey estimated that at present 184,000 people work directly in biotechnology-related industrial activities. It also predicted that overall growth in employment in this area will grow at 1 per cent a year, considerably higher than the general European industry average.

Most of this growth is expected to come from small companies, whose pay-rolls will grow by 6.5 per cent a year. In con-

Two centres, one approach

Two new research centres illustrate the similarities between the way genetics-related research is being carried out in industrial and academic environments.

London. The increasing importance of genetics to the pharmaceutical industry is becoming reflected in a growing convergence not merely between the interests of researchers in the industry and in the academic community, but also in their respective working practices.

Two factors are becoming particularly significant. One is the growing interdisciplinarity of research into deciphering the information contained in the human genome, which — whether in universities or in industry — uses a combination of skills, from genetic analysis through protein engineering to computer science.

The second, related, factor is size. The arrival of genomics — and the enthusiasm for sequencing the hundreds of thousands of nucleotide bases that make up human, animal and plant genomes — has turned biology into a 'big science' for academic and pharmaceutical researchers alike.

This convergence is reflected in the similarities between two new research facilities recently launched in Britain. One is a research centre opened last month by the pharmaceutical company Glaxo Wellcome in Stevenage, 40 miles north of Lon-



Gene analysis at the Sanger Centre.

don. The second is an independent complex currently under development by the Wellcome Trust, with support from the Medical Research Centre and the European Molecular Biology Laboratory, at Hinxton Hall, south of Cambridge.

The two centres have very different functions. The Glaxo centre is intended to become the main focus of the company's research efforts in Britain aimed at producing new drugs. The Hinxton Hall complex, which will include the Sanger Centre and EMBL's European Bioinformatics Institute, remains a basic research facility.

But the Glaxo centre, which will eventually provide high-tech working environment for several hundred scientists, also has something of the air of a university campus, for example through the deliberate inclusion in its design of multiple meeting points (or 'nodes') on the main thoroughfares between laboratories.

"Our goal is to maximize interaction between biologists and chemists, scientists and engineers, or visiting academics and staff researchers," says Alan Baxter, UK research director for the newly formed Glaxo Wellcome. "Our aim has been to provide an environment in which our scientists will spend much more of their time either sitting in front of a personal computer, or at a node discussing new findings in the literature, than in a laboratory."

Hinxton Hall, while engaged in more 'product-oriented' activities than traditional universities laboratories — such as large-scale sequencing and data storage — will similarly strive to maintain a strong feeling as an academic environment.

"In principle there will be more non-PhD scientists than in a traditional institutions, and those working there will have fewer degrees of freedom than they might have at universities," says Michael Morgan of the Wellcome Trust. "But there will still be academic research going on at the same time, and this will be essential to attract top scientists to the different institutions on the campus." **David Dickson**

The boom in bioinformatics

Washington. Employment opportunities in the field of bioinformatics as it relates to genome research have never looked better. A lack of innovative new products in the traditional drug pipeline has sparked considerable interest by many pharmaceutical companies in the new tier of gene-hunting companies, all of whom have made a significant investment in this area. The acute shortage of qualified people largely results from the newness of this field and from a scarcity of formal training programmes that integrate the disciplines of biology and computer science. At least for the foreseeable future, therefore, this is one field where the demand for people with these dual skills will continue to outstrip the supply.

This shortage of qualified people can leave prospective employers in a quandary as to whether to employ a computer scientist who seems fascinated by the biology or a molecular biologist who has been bitten by the computer bug. Mistakes, if they are made, can prove costly. "If the computer system is not done right, at some point that becomes a bottleneck to the overall research," says Tom Slezak, a principal investigator in the informatics core at the Lawrence Livermore National Laboratory Human Genome Center, Livermore, California. "People have to be able to shoot from the hip and hit more often than they miss," he says.

So what are the right set of skills? According to Slezak: "this is a UNIX/C/X-Windows/database type of world."

Training programmes

There is still a "glaring need for training programmes and training opportunities where people at the masters and PhD level would have had formal training in both areas", says Kenneth H. Fasman, informatics director of the Genome Data Base at Johns Hopkins University, Baltimore, Maryland. GDB, as it is called, is a centralized repository for human gene mapping data. "Most of these academic programmes, where they exist at all, are still steaming fresh out of the oven," he says.

One such training programme, run by the University of Pennsylvania, is attempting to bridge that gap and provide researchers with a solid grounding in both the computational side and the molecular biology. The Penn Computational Biology Training Program is only about six months old and offers interdisciplinary training at both the graduate and post-doctoral level with funding from the US National Science Foundation. Funding is available for four graduate students and two post-docs at steady state, says Chris Overton, research associate professor in depart-

To take full advantage of the vast wealth of genetic data emanating from the US government-funded Human Genome Project, and other independent efforts, new software tools and databases must be developed for the collection, ordering and interpretation of the data, so that it can then be used for product development.

ments of genetics and of computer and information sciences, and one of 21 faculty taking part in the programme. In addition, both Merck and SmithKline Beecham recognize the need for qualified people with these dual skills and have agreed to provide support for the training of one post-doc each.

Graduates enter the programme through the biology, genetics or computer and information science departments, but are then given additional training in the complementary discipline. Particular areas of interest include physical and genetic mapping, biological databases, multiple sequence alignment, phylogeny construction, sequence search and analysis, statistical methods, and discrete algorithms and combinatorial optimization in biology. Training will also be offered at the post-doctoral level for researchers with interdisciplinary research interests. (For more information contact: <http://www.cbil.upenn.edu/~dsearls/NSF.html>)

Opportunity knocks

As a discipline, bioinformatics is critical to the success of the smaller gene-hunting companies like InCyte Pharmaceuticals (Palo Alto, California) and Human Genome Sciences (HGS) (Gaithersburg, Maryland), the commercial arm of a partnership that also includes The Institute for Genomic Research. Both of these companies operate high-throughput cDNA sequencing efforts. Companies like Myriad Genetics (Salt Lake City, Utah), who last year isolated the elusive *BRCA1* breast and ovarian cancer gene, Mercator Genetics (Menlo Park, California), Millennium Pharmaceuticals (Cambridge, Massachusetts) and Sequana Therapeutics (La Jolla, California) are taking a different approach. They, instead, are hoping to identify human genes associated with a specific disease by looking at large numbers of families in which that disease is particularly prevalent.

Carlos Zamudio, director of bioinformatics and engineering at Sequana, which is focusing on some of the more complex diseases like asthma, diabetes, obesity and osteoporosis, says Sequana has made a heavy investment in bioinformatics. This area now makes up roughly

one-third of its research effort but Zamudio says this could grow to 50 per cent as the company moves from finding genes to characterizing genes. Ultimately that will mean training people in-house, he says.

Although slower off the mark, many of the larger pharmaceutical companies are now seeing real value in gene mapping and sequence data. They are either showering money around among the smaller gene-hunting companies, or are setting up bioinformatics advisory boards and attempting to lure experts away from academia. One of the first to jump on the human genome bandwagon was SmithKline Beecham who, in 1993, made a \$125-million investment in HGS. Others have subsequently stepped up to the plate, including Merck (which is funding a gene sequencing effort at Washington University), Pfizer (which has linked up with InCyte) and Hoffmann-La Roche (which has done a deal with Millennium).

Salaries

The job opportunities for someone with the right skills may look rosy, but the money isn't bad either. The reason: simple supply-and-demand economics. Someone with a masters in computer science can typically command more than an equivalent-level molecular biologist, and probably more than someone at the PhD level. As Slezak puts it: "Good computer people cost a lot more than good biologists, bad computer people cost more than good biologists." Zamudio says he has employed some people where the supply is so low that basically they can command their own salary. "Mostly, they're not experienced enough to know how far they can take it," he quips.

Networking

Several upcoming meetings might be of interest: (1) The 4th International Conference on Bioinformatics & Genome Research, San Francisco, California, USA, 5-7 June '95, sponsored by Cambridge Healthtech Institute (tel. +1 617-487-7989 or <http://id.wing.net/~chi/homepg.html> via the WWW); and (2) two back-to-back meetings in Cambridge, UK — Intelligent Systems in Molecular Biology, 16-19 July '95 (contact Dominic Clark at ICRF, London, dac@bison.lif.icnet.uk or <ftp://ftp.icnet.uk/icrf-public/ismb/ismb95.html> via WWW) and the Second Meeting on the Interconnection of Molecular Biology Databases, 20-22 July '95 (contact Peter Karp at the Stanford Research Institute, California, pkarp@ai.sri.com or <http://www.ai.sri.com/people/pkarp/mimbd.html> via WWW).

Diane Gershon